



Universidade de Brasília

Instituto de Ciências Exatas  
Departamento de Ciência da Computação

## **Análise de Dados Abertos sobre o Ensino Superior Brasileiro**

Halley da Fraga Magalhães  
Luan de Abreu Cardoso

Monografia apresentada como requisito parcial  
para conclusão do Curso de Computação — Licenciatura

Orientadora  
Prof.a Dr.a Maristela Terto de Holanda

Coorientador  
Prof. Dr. Marcio de Carvalho Victorino

Brasília  
2016



# Análise de Dados Abertos sobre o Ensino Superior Brasileiro

Monografia apresentada como requisito parcial  
para conclusão do Curso de Computação — Licenciatura

Prof.a Dr.a Maristela Terto de Holanda    Prof. Dr. Marcio de Carvalho Victorino  
CIC/UNB    CIC/UNB

Prof. Dr. Edison Ishikawa  
CIC/UNB

Prof. Dr. Pedro Antonio Dourado de Rezende  
Coordenador do Curso de Computação — Licenciatura

Brasília, 09 de dezembro de 2016

# Dedicatória

A Deus pela força que nos deu em tempos de escuridão, às nossa famílias que acreditaram que a educação faz a diferença e aos amigos que auxiliaram na caminhada.

# Agradecimentos

Agredecemos à Deus pela oportunidade dada e aos familiares que tanto nos apoiaram em garantir uma educação de qualidade e acesso ao ensino superior que não tiveram.

# Resumo

Existe uma grande quantidade de dados gerada atualmente, principalmente, quando falamos em nível de instituições governamentais. Essa projeção se mostra consideravelmente mais ampla, tornando o processo de tomada de decisões cada vez mais dificultoso e com a necessidade de uma orientação mais técnica e factual. Processo esse que demanda a construção de um banco de dados analítico - DW, para geração de relatórios analíticos.

Este trabalho tem por objetivo construir um modelo de Análise de Dados Abertos sobre o Ensino Superior Brasileiro, por meio de uma *Data Warehouse*, com dados disponibilizados pelo Instituto Nacional de Estudos e Pesquisa Anísio Teixeira (INEP), para a geração de relatórios analíticos (*Dashboards*) no intuito de auxiliar no processo de tomada de decisão.

**Palavras-chave:** Dados Abertos, Ensino Superior, SAD, DW, Dashboards, INEP

# Abstract

There is a large amount of data currently generated, especially when speak level of government institutions, this projection is shown considerably wider making the process of making decisions increasingly difficult and with the need for more technical and factual orientation. A process that requires the construction of an analytical database - Data Warehouse to generate analytical reports.

This work aims to build an Open Data Analysis model on the Brazilian Higher Education, through a Data Warehouse with data provided by the National Institute of Studies and Research Anísio Teixeira (INEP), for generating analytical reports (dashboards) in order to assist in the decision-making process.

**Keywords:** Open Data, Higher Education , SAD, DW, Dashboards, INEP

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Definição do Problema . . . . .	2
1.2	Hipótese . . . . .	2
1.3	Motivação . . . . .	2
1.4	Objetivos . . . . .	2
1.4.1	Objetivo Geral . . . . .	3
1.4.2	Objetivos Específicos . . . . .	3
1.5	Estrutura do trabalho . . . . .	3
<b>2</b>	<b>Dados Abertos</b>	<b>4</b>
2.1	Dados . . . . .	4
2.1.1	Dados e Informações . . . . .	4
2.1.2	Dados Abertos . . . . .	5
2.1.3	Padronização por Recurso Identificador Uniforme - Uniform Resource Identifier (URI) . . . . .	7
2.1.4	Dados Governamentais . . . . .	8
<b>3</b>	<b>Sistemas de Informação</b>	<b>10</b>
3.1	Sistema de Informação . . . . .	10
3.1.1	Definição de Sistema de Informação . . . . .	10
3.1.2	Classificação dos Sistemas de Informação . . . . .	11
3.2	Processo de Tomada de decisão e gestão do conhecimento . . . . .	12
3.2.1	Sistema de Apoio à Decisão (SAD) . . . . .	13
3.3	Conceituação de <i>Data Warehouse</i> (DW) . . . . .	14
3.3.1	Introdução à Modelagem Dimensional . . . . .	17
3.3.2	Sistemas OLTP e OLAP . . . . .	18
3.3.3	Esquema Estrela comparado aos Cubos OLAP . . . . .	19
3.3.4	Processo de Extração, Transformação e Carga - ETL . . . . .	21

<b>4</b>	<b>Estudo de Caso</b>	<b>22</b>
4.1	Do Objeto de estudo . . . . .	22
4.2	Dados obtidos sobre a Educação Superior Brasileira . . . . .	22
4.3	Histórico da Instituição - INEP . . . . .	23
4.3.1	Instituto Nacional de Estudos e Pesquisa Anísio Teixeira (INEP) . . .	23
4.3.2	Sobre o INEP . . . . .	24
4.3.3	Processo de Downloads de Microdados do Ensino Superior Brasileiro .	25
<b>5</b>	<b>Implementação do Estudo de Caso</b>	<b>27</b>
5.1	Implementação do Estudo de Caso - Análise de Dados Abertos do Ensino Superior Brasileiro . . . . .	27
5.1.1	Implementação do processo de Extração, Transformação e Carga - ETL	30
5.1.2	Processo de Extração: . . . . .	31
5.2	Processo de Transformação: . . . . .	32
5.3	Cargas realizadas na STAGE para serem carregadas nas tabelas dimensões .	33
<b>6</b>	<b>Resultados</b>	<b>34</b>
6.1	Resultados Obtidos após a realização de consultas . . . . .	34
<b>7</b>	<b>Conclusões e Trabalhos Futuros</b>	<b>42</b>
7.1	Considerações Finais . . . . .	42
7.2	Trabalhos Futuros . . . . .	42
	<b>Referências</b>	<b>44</b>
	<b>Apêndice</b>	<b>45</b>
<b>A</b>	<b><i>Dashboards</i>, criação de tabelas, diagrama ER, formulários de preenchimento</b>	<b>46</b>



# Lista de Abreviaturas e Siglas

**ASCII** *American Standard Code for Information Interchange.*

**BD** Banco de Dados.

**BI** *Business Intelligence.*

**CSV** *Comma-separated values.*

**DBA** *Database Administrator.*

**DW** *Data Warehouse.*

**ENCCEJA** Exame Nacional Para Certificação de Competências.

**ENEM** Exame Nacional do Ensino Médio.

**ER** Entidade-Relacionamento.

**ETL** *Extract, Transform, Load.*

**FIES** Programa de Financiamento Estudantil.

**GPL** *General Public License.*

**IBGE** Instituto Brasileiro de Geografia e Estatística.

**IES** Instituições de Ensino Superior.

**INEP** Instituto Nacional de Estudos e Pesquisa Anísio Teixeira.

**MEC** Ministério da Educação.

**OLAP** *On-Line Analytic Processing.*

**OLTP** *On-line Transaction Processing.*

**PRONATEC** Programa Nacional de Acesso ao Ensino Técnico e Emprego.

**RDF** *Resource Description Framework*.

**SAD** Sistema de Apoio à Decisão.

**SAE** Sistemas de Apoio Executivo.

**SAEB** Sistema Nacional de Avaliação da Educação Básica.

**SGBD** Sistemas Gerenciadores de de Banco de Dados.

**SI** Sistemas de Informação.

**SIG** Sistemas de Informação Gerenciais.

**SINAES** Sistema Nacional de Avaliação da Educação Superior.

**SPT** Sistemas de Processamento de Transações.

**SQL** *Structured Query Language*.

**STC** Sistemas Trabalhadores do Conhecimento.

**TCU** Tribunal de Contas da União.

**TXT** *Text File*.

**URI** Uniform Resource Identifier.

**URL** Uniform Resource Locator.

**UTF-8** *8-bit Unicode Transformation Format*.

**W3C** World Wide Web.

# Capítulo 1

## Introdução

O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - INEP, tem por missão a promoção de estudos, pesquisas e avaliações sobre o Sistema Educacional Brasileiro com o objetivo de subsidiar a formulação e implementação de políticas públicas para a área educacional a partir de parâmetros de qualidade e equidade, bem como produzir informações claras e confiáveis aos gestores, pesquisadores, educadores e público em geral [7].

Os dados do INEP são disponibilizados e podem ser extraídos por meio de um sítio na internet [portal.inep.gov.br](http://portal.inep.gov.br) de maneira aberta sem restrição de acesso para qualquer pessoa poder livremente usá-los, reutilizá-los e redistribuí-los [5].

Este trabalho tem por objetivo realizar a análise de dados abertos do Ensino Superior Brasileiro obtidos pelo sítio do INEP.

A análise de dados abertos foi feita mediante o processamento analítico *On-Line* ( *On-Line Analytic Processing* (OLAP)) que é constituído de todas atividades gerais de consulta a apresentação de dados numéricos e textos provenientes de um *Data Warehouse*, assim como as formas específicas de consultas e apresentações que são exemplificadas por uma grande quantidade de ferramentas OLA, apresentados por Kimball *et al.* [10].

Os dados disponibilizados pelo INEP foram extraídos, transformados e carregados em um DW, com o objetivo de auxiliar mediante um Sistema de Apoio de Decisões - SAD, para a realização de análises do ensino superior brasileiro.

## 1.1 Definição do Problema

O INEP disponibiliza os dados sobre o ensino superior brasileiro no respectivo sítio na web [6] em arquivos compactados, extensão '.zip', entretanto, a falta de integração e a descrição heterogênea da semântica dos dados dificultam sua análise de forma integrada e sumarizada.

Não existe um banco de dados analítico público de fácil acesso, na qual tanto a sociedade como gestores tenham acesso aos dados de maneira rápida com agregação de dados e geração de relatórios.

## 1.2 Hipótese

Acredita-se que após um processo de *Extract, Transform, Load* (ETL) - desses dados em um *Data Warehouse*, o desenvolvimento de relatórios OLAP e a análise facilitarão e proporcionarão *insights* mais significativos.

## 1.3 Motivação

A análise de dados abertos pode proporcionar vários benefícios, dentre eles:

- a melhoria dos serviços públicos e uma melhor compreensão das atividades governamentais;
- gestão mais efetiva dos serviços públicos;
- aumento da responsabilização e da prestação de contas dos serviços públicos prestados à sociedade;
- aumento da integridade e reputação pública; e
- maior participação do cidadão na gestão pública.

Então, tem-se neste contexto a abordagem para o tratamento dos dados disponibilizados por organizações públicas na web, a fim de facilitar a análise, tornando-a de fundamental importância.

## 1.4 Objetivos

Este trabalho está dividido em Objetivo Geral e Objetivos Específicos.

### 1.4.1 Objetivo Geral

O objetivo geral deste trabalho é executar o processo de extração e transformação dos dados disponibilizados pelo INEP na *web*, para posterior carga em um *Data Warehouse* para realização de consultas analíticas.

### 1.4.2 Objetivos Específicos

Para alcançar o objetivo geral deste trabalho, os seguintes objetivos específicos foram necessários:

- Criar um modelo relacional compatível com o domínio observado.
- Executar a extração, transformação para posterior carga desses dados no sistema gerenciador de banco de dados relacional *MYSQL* modelado de forma multidimensional.
- Desenvolver relatórios analíticos utilizando a ferramenta *OpenSource - Pentaho*.
- Executar consultas ao *Data Warehouse* utilizando os relatórios analíticos desenvolvidos na ferramenta *Pentaho* para apresentar análises sobre o ensino superior brasileiro.

## 1.5 Estrutura do trabalho

Para atingir os objetivos propostos por esse trabalho, estrutura-se da seguinte maneira:

- Capítulo 2: Fundamentação teórica acerca da definição de dados, dados abertos, informação e dados governamentais.
- Capítulo 3: Definição e fundamentação teórica de Sistemas de Informação, Sistemas de Apoio à Decisão, modelagem dimensional, ciclo de vida de um projeto de um armazém de dados (DW), bem como sua arquitetura.
- Capítulo 4: Estudo de caso em uma aplicação de DW com uso de dados abertos.
- Capítulo 5: Implementação do estudo de caso para mostrar a obtenção de relatórios analíticos mediante a construção de um *Data Warehouse*.
- Capítulo 6: Geração de *Dashboards* e relatórios analíticos.
- Capítulo 7: Conclusões sobre a geração dos relatórios analíticos.

# Capítulo 2

## Dados Abertos

Neste Capítulo são apresentados fundamentais para a elaboração deste trabalho. Dentre esses, na Seção 2.1 é apresentado o conceito de Dados, com as respectivas Subseções 2.1.1 Dados e Informações; 2.1.2 Dados Abertos; 2.1.3 Padronização de Recursos; 2.1.4 Dados Governamentais.

### 2.1 Dados

#### 2.1.1 Dados e Informações

O ambiente informacional atual é composto por um enorme volume de dados e informações. Machado [13] afirma que a informação acrescenta algo ao conhecimento da realidade a ser analisada, como, por exemplo, a dosagem de um determinado medicamento que um paciente receber, pode ser entendida como uma informação.

Esse conhecimento pode ou não ser modelado (registrado), já que o Dado, conforme Machado [13], representa, um registro de uma informação. Este Dado pode ser registrado fisicamente através de um papel (receita médica), um disco de computador ou impulso elétrico, etc. Este registro pode ser o originador de uma série de processos que influenciam a realidade observada, como o salvamento de paciente, o sinal de alarme e outros.

Ainda conforme Machado [13], o tratamento das Informações dá origem a vários tipos de Dados, porém o dado deve registrar apenas os aspectos realmente relevantes da informação, ou seja, o endereço do fabricante do remédio não tem nenhum interesse para um sistema de controle que mantém a vida dos pacientes no hospital. Pode-se concluir, então, que em um sistema de informação, estão contidas todas informações necessárias ao

objetivo do sistema (de manter a vida do paciente). Os dados originados dessas informações serão processados pelo sistema criado. Por definição, um computador não processa informações, mas sim dados.

Por outro lado, Rob e Coronel [18], afirmam que os dados são fatos brutos, sendo bruto o fator de que os fatos ainda não foram devidamente processados para revelar seu real significado, como um formulário de entrevista web para critérios de satisfação e que ao término de colhimento de resposta terá apenas os fatos, mas não a síntese do que significa esses fatos.

Já informação, segundo próprio Rob e Coronel [18] se dá pelo resultado de processamento de dados brutos para que possam revelar seu real significado. Nesses procedimentos estão uma simples organização de dados de modo a revelar padrões, realização de previsões ou extrações de inferência por meio de estatística, além da devida contextualização da informação.

Tem-se também que informação conforme, Laudon e Laudon [12], representa dados apresentados em uma forma significativa e útil para seres humanos, enquanto que dados, ao contrário, são sequências de fatos ainda não analisados, representativos, de eventos que ocorrem de fato. Logo, temos que dados são algo armazenado em um banco de dados e que as informações são os significados obtidos por esses dados para um determinado usuário, grupo de usuários, instituição ou organização.

Dados tem como as seguintes espécies e variantes, tais quais : dados abertos e governamentais.

### 2.1.2 Dados Abertos

Dados abertos, conforme sítio <http://dados.gov.br/dados-abertos> [5], possuem as seguintes características:

- **Aberto:** significa que qualquer um pode livremente ter acesso, uso, modificação, compartilhamento e para qualquer finalidade, podendo estar sujeito, no máximo, às exigências que preservem a proveniência - origem do dado - e sua abertura.
- **Dados Abertos e Conteúdo Aberto:** pode ser livremente usado, modificado e compartilhado por qualquer pessoa e para qualquer finalidade.

A Open Knowledge Foundation (Fundação de Conhecimento Aberto) [2] define o termo Dados Abertos em três formas fundamentais:

- **Disponibilidade e Acesso:** os dados devem estar disponíveis como um todo e sob custo não maior que um custo razoável de reprodução, e que possam haver *download* pela Internet. Estes dados devem estar disponíveis de uma forma conveniente e modificável.
- **Reúso e Redistribuição:** os dados devem ser fornecidos sob termos e licenças que possam ser utilizados e redistribuídos, inclusive com a combinação com outros conjuntos de dados.
- **Participação Universal:** todos devem ser capazes de usar, reutilizar e redistribuir - não deve haver discriminação contra áreas de atuação ou contra pessoas ou grupos. Restrições comerciais e educacionais estariam nesse rol discriminatório.

A disponibilização e abertura de dados abertos tem como preceito seguir uma classificação de 5 estrelas, afirmado por Isotani e Bittencourt [9], elencados em:

- ★ Disponível na Internet (em qualquer formato; ODT,PDF), desde que possua licença aberta, para que seja considerado aberto.
- ★★ Disponível na Internet de maneira estruturada (em um arquivo CALC com extensão ODF).
- ★★★ Disponível na Internet, de maneira estruturada e em formato não proprietário (CSV em vez de Excel).
- ★★★★ Seguindo os itens anteriores, mas dentro dos padrões estabelecidos pelo World Wide Web (W3C) (tais quais *Resource Description Framework* (RDF) e *SARQL*): usar Uniform Resource Locator (URL) para identificar coisas e propriedades, de modo a fazer com que os usuários possam direcionar suas publicações.
- ★★★★★ Todos os itens anteriores, e conectar seus dados a outros dados, de forma a fornecer um contexto ou situação.

Para publicação de Dados Abertos foram definidos dez boas práticas, conforme Consortium et al. [1]:

1. **Preparar os stakeholders:** esta etapa é voltada para a formação dos usuários que criarão e manterão os Dados Abertos Conectados.
2. **Selecionar a fonte dos dados:** etapa dedicada a definir que se pretende abrir e conectar a outros dados e disponibilizar ele para o reúso.



3. **Modelar os dados:** com as partes interessadas capacitadas e o conjunto de dados definidos, começa a etapa de modelagem dos Dados Conectados. Ou seja, como serão apresentados os dados e como eles se relacionam com outros dados de forma independente de aplicação.
4. **Especificar a licença:** nesta etapa, a organização, responsável pelos dados que serão abertos, deve definir e especificar a licença que será usada.
5. **Nomear bons Uniform Resource Identifier (URI)s:** esta etapa é o núcleo dos Dados Abertos Conectados, e a definição e o uso de boas práticas para Uniform Resource Identifier (URI)s é fundamental.
6. **Usar vocabulários padronizados:** uma das melhores formas de conectar dados é por meio do reuso de vocabulários conhecidos, como as recomendações do World Wide Web (W3C).
7. **Converter os dados:** uma vez que a estratégia de modelagem e as boas práticas para Uniform Resource Identifier (URI)s foram definidas e os vocabulários a serem reutilizados forem identificados, vem a etapa de converter os dados da fonte original para representação adequada aos Dados Conectados.
8. **Prover acesso aos dados:** esta etapa define quais serão as formas de acesso que usuários e máquinas terão aos dados.
9. **Anunciar novo conjunto de Dados Conectados:** de nada adianta conectar os dados e não anunciar para a sociedade que eles foram disponibilizados. Desta forma, esta cumpre este papel de divulgação de um novo conjunto de dados (*Dataset*) publicado.
10. **Reconhecer a função social:** esta etapa é definida para que o responsável por publicação dos dados cumpra a função de manter os dados publicados ao longo do tempo.

### 2.1.3 Padronização por Recurso Identificador Uniforme - Uniform Resource Identifier (URI)

Com a padronização de Recursos Identificadores Uniformes - URIs, que se trata de uma cadeia de caracteres utilizados com o fim de identificar um determinado recurso mediante padrões de uso previamente definidos, Isotani e Bittencourt [9], afirmam que é possível com uso de URIs:

- a) Conectar e combinar dados com outros dados;
- b) Reusar dados em outros contexto;
- c) Melhorar a busca e a compreensão dos dados apresentados;
- d) Possibilitar inferência a partir de dados parciais;
- e) Permitir navegação entre documentos.

#### 2.1.4 Dados Governamentais

David Eaves [3], especialista em políticas públicas e defensor dos dados abertos definiu as seguintes premissas para dados abertos:

- **Completos:** todos os dados públicos são disponibilizados. Dados são informações eletronicamente gravadas, incluindo, mas não sendo limitadas, somente a documentos, bancos de dados, transcrições e formatos audiovisuais. Dados públicos são dados que não estão sujeitos a limitações válidas de privacidade, segurança ou controle de acesso reguladas por estatutos.
- **Primários:** os dados são publicados na forma coletada na fonte, com mais fina granularidade possível, e não de forma agregada ou transformada.
- **Atuais:** Os dados são disponibilizados o mais recente possível para preservar seu valor e tempestividade.
- **Acessíveis:** os dados são disponibilizados para o público mais amplo possível e para propósitos quaisquer.
- **Processáveis por máquina:** os dados são razoavelmente estruturados para possibilitar o seu processamento automatizado.
- **Acesso não discriminatório:** os dados estão disponíveis a todos, não havendo identificação ou registro necessário.
- **Formatos não proprietários:** os dados estão disponíveis em um formato em que não haja nenhum controle exclusivo ou direito de propriedade única.
- **Livres de licenças:** os dados não estão sujeitos a regulações de direitos autorais, marcas patentes ou segredo industrial. Restrições razoáveis de privacidade, segurança e controle de acesso serão permitidas quando reguladas por instruções normativas.

No entanto, tais premissas ou leis, como conceitua David Eaves, apesar de propostas para dados governamentais, também podem se dar para todos os dados abertos de forma geral.

Adotado pelo Governo Federal Brasileiro e citado no sítio <http://dados.gov.br> [5], o estabelecimento de dados governamentais ocorreu em 2007, na qual um grupo de trabalho de 30 pesquisadores se reuniram na Califórnia, Estados Unidos Da América, para definir os princípios dos Dados Abertos e em seguida as três leis apontadas por David Eaves [3]:

1. Se o dado não pode ser encontrado e indexado na Web, ele não existe;
2. Se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado; e
3. Se algum dispositivo legal não permitir sua replicação, ele não é útil.

A Administração Pública Federal Brasileira busca a cada dia se mostrar transparente dado o princípio constitucional da publicidade de seus atos e atuações diante da sociedade brasileira, com a promoção de políticas e investimentos de dados abertos para maior controle e participação social - processo denominado de *Accountability* - que em livre tradução seria a prestação de contas, responsabilização.

Esse processo de responsabilização permite a avaliação pela sociedade para analisar as ações e decisões governamentais e cobrar devidas medidas necessárias, além da fiscalização da atividade e gerência pública, processo intimamente ligado a conceitos de democracia.

A transmissão de dados abertos auxiliam nos seguintes aspectos:

1. Realização de novos estudos, no aspecto de inovação;
2. Realização de atividades e usos pelo próprio Governo e gerador desses dados, tendo como exemplo o INEP;
3. Realização de estudos pelos cidadãos, sejam eles estudantes universitários, pesquisadores, programadores, desenvolvedores, professores e outros.

O que configura um auxílio externo para se aprimorar a qualidade e os estudos dos dados governamentais, explicitados pelo TCU [20].

# Capítulo 3

## Sistemas de Informação

### 3.1 Sistema de Informação

Neste capítulo são apresentados a definição de Sistemas de Informação e sua devida classificação, bem como o processo de tomada de decisões, apoio a decisão, além da conceituação de *Data Warehouse*.

#### 3.1.1 Definição de Sistema de Informação

Segundo Laudon e Laudon [12] um Sistema de Informação (SI) pode ser definido como um conjunto de componentes inter-relacionados que coletam (ou recuperam), processam, armazenam e distribuem informações destinadas a apoiar a tomada de decisões, a coordenação e o controle, esses sistemas também auxiliam os gerentes e trabalhadores a analisar problemas, visualizar assuntos complexos e criar produtos ou serviços. De acordo com Laudon e Laudon, um SI compreende:

- A entrada que consiste na captura ou coleta de dados brutos de dentro da organização ou de seu ambiente externo;
- O processamento que consiste na conversão de dados brutos em uma forma mais significativa;
- A saída que consiste na transferência de informações processadas às pessoas que farão uso ou às atividades nas quais serão empregadas essas informações. Também faz-se necessário que os sistemas de informações tenham uma retroalimentação desse sistema de modo a dar uma resposta à ação que foi utilizada para determinados funcionários ou usuários da organização de tal maneira que possa corrigir erros ou avaliá-los.

### 3.1.2 Classificação dos Sistemas de Informação

Laudon e Laudon [12] explicam de maneira clara e objetiva que Sistemas de Informação podem ser organizados em uma classificação de 4 níveis ou estratos, são eles: operacional, conhecimento, gerencial e estratégico. Esses sistemas são divididos em:

**Sistemas de Processamento de Transações (SPT)** correspondem a sistemas de menor nível hierárquico em uma organização, atendendo as necessidades operacionais e informações armazenadas de maneira estruturada, como em tabelas. São usados por técnicos e especialistas que possuem função de executar e cumprir os planos e metas estabelecidos pelos outros sistemas e servem como base na entrada de dados. São atividades intensamente estruturadas e corriqueiras.

**Sistemas Trabalhadores do Conhecimento (STC)** é um SI de nível organizacional que recebe dados operacionais, onde decodifica e cria novos conhecimentos que atende as necessidades de informação da organização, são apresentadas por um grupo de especialistas e analistas da organização.

**Sistemas de Informação Gerenciais (SIG)** visam a atender o nível gerencial da organização, de modo a prover relatórios gerenciais e analíticos, e sendo o acesso, normalmente, imediato (*on-line*) às ocorrências de desempenho e a dados históricos. Possuem orientação a fatos e ocorrências internas, não se preocupando em visualizar situações externas à organização. Habitualmente dependem dos sistemas transacionais próximos para obtenção de dados e cabe a eles o processo de sumarização de dados e a emissão de relatórios estabilizados sobre as operações da organização.

**Sistema de Apoio à Decisão (SAD)** atendem ao nível gerencial da organização, auxiliando no processo de tomada de decisões não corriqueiras e usuais que tendem a alterar com rapidez e que não são facilmente especificadas com precedência. SADs utilizam as informações internas geradas pelo SPT e pelo SIG e também fazem uso de informações de fontes externas que sejam de interesse da organização. Nesse sistema temos a ocorrência do processo de ETL para adiante serem armazenadas em um repositório dimensional, conhecido por *Data Warehouse*.

**Sistemas de Apoio Executivo (SAE)** atende ao nível estratégico da organização, que é utilizado pela alta cúpula para tomada de decisões. SAEs abordam decisões não usuais que exigem bom senso, avaliação e percepção. Oferecem um ambiente difuso de computação, em lugar de oferecer aplicação específica. Incorporam dados sobre eventos e ocorrências externas em conjunto a informações resumidas do SIG e SAD

Há uma grande necessidade da realização de troca de informações entre os sistemas de informações dos mais diversos níveis e estratos de uma organização. Visto na figura abaixo:

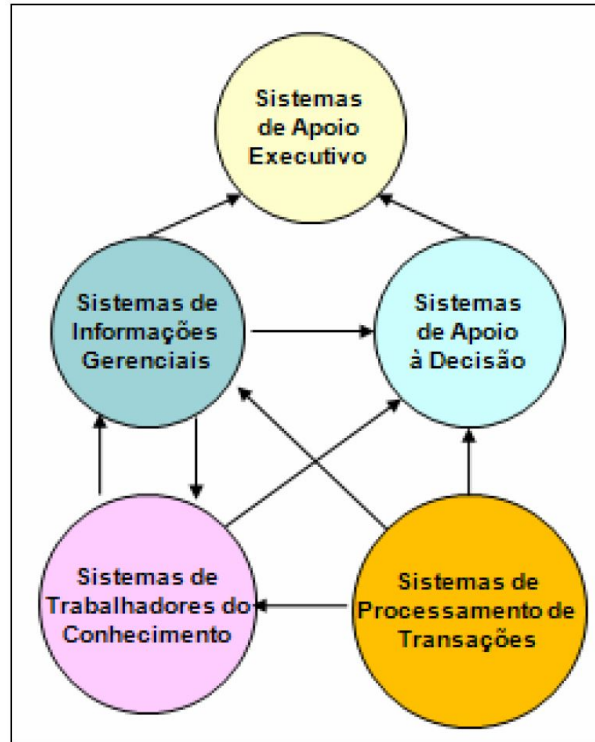


Figura 3.1: Relacionamentos entre os sistemas de informações em uma organização [12].

Os sistemas de apoio à decisão armazenam os dados oriundos dos mais variados sistemas, como sistemas de processamento de transações, sistemas de informações gerenciais, sistemas de trabalhadores do conhecimento e também fonte de dados externos como no inter-relacionamento da Figura 3.1. Esse armazenamento se dá em armazéns ou repositórios multidimensionais chamados de *Data Warehouse* para geração de relatórios analíticos.

## 3.2 Processo de Tomada de decisão e gestão do conhecimento

Uma das principais contribuições dos sistemas de informação é a melhoria da tomada de decisão, seja para indivíduos ou grupos. A tomada de decisão nas empresas costumava limitar-se à diretoria. Contudo, atualmente, funcionários de níveis mais baixos são responsáveis por algumas dessas decisões, na medida em que os sistemas de informação tornam

os dados disponíveis para camadas mais elementares da empresa. E consoante Laudon e Laudon [12] possuem três classificações para o processo de tomada de decisão:

**Decisões não estruturadas:** que são aquelas em que o responsável pela tomada de decisão faz uso do bom-senso, da experiência e da capacidade de avaliação e sua visão na definição do problema. Cada uma dessas decisões é inconstante, importante e não rotineira, sem que o devido procedimento seja bem compreendido e definido.

**Decisões estruturadas:** são repetitivas, rotineiras e definidas e envolvem procedimentos predefinidos, de modo que não precisam ser tratadas como se fossem novas.

**Decisões Semiestruturadas:** apenas parte do problema tem uma resposta clara e definida, dada por um procedimento aceito. Tem-se que que decisões estruturadas são mais corriqueiras nos níveis organizacionais mais baixos, enquanto problemas não estruturados são relativos a alta administração da organização que tem metas para serem traçadas nos próximos anos.

### 3.2.1 Sistema de Apoio à Decisão (SAD)

O processo de elaboração desse projeto de Análise de Dados Abertos do Ensino Superior Brasileiro segue o Sistema de Apoio à Decisão para dar apoio à análise de problemas semiestruturados e não estruturados.

Laudon e Laudon [12] afirmam que a capacidade do SAD se dá pela análise baseada em uma teoria ou modelo bem fundamentado que quando combinados com uma boa interface de usuário, tornam o sistema simples de usar. Alguns SADs atualmente são orientados por dados, usando processamento analítico on-line e mineração de dados - *Data Mining* para analisarem enormes quantidades de dados em grandes sistemas corporativos ou base de dados abertos.

E afirmam também que aplicações de *Business Intelligence* - BI quando orientadas por dados apoiam a tomada de decisão por permitir que os usuários extraiam informações úteis escondidas em grandes quantidades de dados.

A Figura 3.2 mostra os principais componentes de um Sistema de Apoio à Decisão que são o banco de dados SAD, no caso desse projeto, o banco de dados analítico ou repositório de dados (*Data Warehouse*), o *software* de sistema SAD, no caso o uso do *Pentaho Data-Integration* e o *Pentaho BI-server* para interface junto ao usuário.

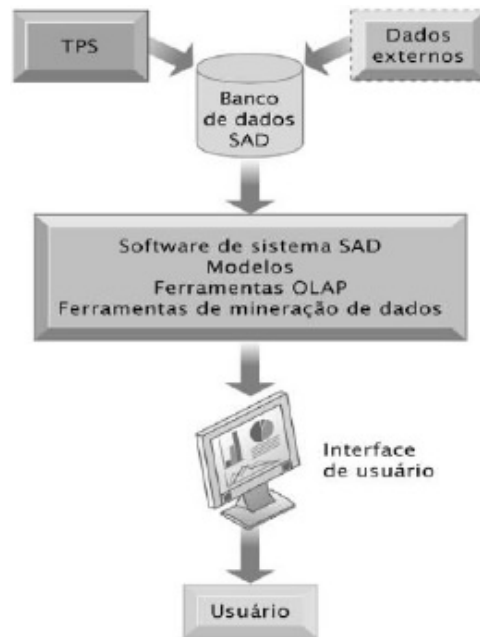


Figura 3.2: Visão geral de um Sistema de Apoio à Decisão. Conforme: Laudon e Laudon [12].

O banco de dados SAD é uma coletânea de dados correntes ou históricos, que em nosso projeto se trata da Análise do Ensino Superior Brasileiro durante o decorrer dos anos, proveniente de uma série de aplicações ou grupo. Em se tratando desse projeto, o banco de dados SAD é um *Data Warehouse* com o objetivo de ser atualizado de forma contínua por entradas de dados de novos anos.

Para a análise de Dados, conforme [12], tem-se o uso de ferramentas OLAP, ferramentas de mineração de dados ou um conjunto de modelos matemáticos e analíticos que podem ser disponibilizados para o usuário do SAD. Além da visualização de dados que ajudam os usuários a perceber padrões e relações em grandes quantidades de dados, o que seria difícil visualizar em listas de texto tradicionais.

### 3.3 Conceituação de *Data Warehouse* (DW)

*Data Warehouse* conforme Poe *et al.* [16] é um banco de dados analítico usado como base para um sistema de suporte à decisão.



Inmon [8] define DW como uma coleção de dados orientadas por assuntos, integrada, não volátil e variável em relação ao tempo, que tem por objetivo dar suporte aos processos de tomada de decisões.

Gardner [4] menciona DW como um processo, não um produto ou serviço, para montagem e administração de dados provenientes de várias fontes com o propósito de obter uma simples e detalhada visão de parte de todo o negócio.

Sen e Jacob [19] apontam DW como uma construção no interesse de dar suporte à decisão de negócios, bem como conter dados históricos sumarizados e consolidados procedente de registros individuais de banco de dados operacionais.

Corey *et al.* [14] explicitam DW como uma coleção de informações corporativas, derivadas diretamente de sistemas operacionais e algumas fontes externas. Tendo o propósito específico de suportar decisões de negócios e não operações de negócios.

Klein *et al.* [11] consideram que um *Data Warehouse* proveem de um banco de dados especializado capaz de manipular uma grande quantidade de informações com um bom desempenho em seu acesso, dado a melhora da gerência, do seu controle e do acesso a esses dados. Dados esses obtidos por dados operacionais ou fontes externas extraídos através de ferramentas específicas, ou obtidos de outras fontes, na qual deverão ser tratados e integrados em um banco de dados orientados por assuntos.

E ainda conforme Klein *et al.* [11] as ferramentas destinadas à etapa de transformação dos dados em conhecimento e informação, entre elas, as ferramentas *On-Line Analytic Processing* (OLAP), que oferecem ao usuário uma elevada capacidade de manipulação, permitindo que sejam analisados os resultados obtidos.

Logo DW é um repositório ou armazém de dados de uma organização para geração de relatórios analíticos para suporte ao processo de tomada de decisões.

Muitas vezes estruturas normalizadas em um sistema de Banco de Dados são considerados pela indústria, instituições, empresas e organizações como um *Data Warehouse* empresarial; no entanto, Kimball *et al.* [10] citam que estas terminologias se tornam um equívoco, pois o armazém de dados é considerado muito mais abrangente do que um conjunto de tabelas normalizadas.

*Data Warehouse* empresarial refere-se mais precisamente ao conglomerado de dados

de uma organização através do armazenamento, preparo e apresentação de informações. Assim, diante dessas premissas tem-se um *Data Warehouse* empresarial, institucional ou organizacional como a união de todos os diversos componentes, sistemas e informações contidas em um armazenamento de dados [10].

Um dos fatores cruciais que Kimball *et al.* [10] citam é que algumas equipes de *Database Administrator* (DBA) projetistas de DW falham consideravelmente, porque eles concentraram toda a sua energia e recursos na construção de estruturas normalizadas em vez de alocação de tempo para o desenvolvimento de uma área de apresentação que apoia de maneira mais eficaz a tomada de decisão de negócios.

Então, de modo a realizar uma construção de relatórios e pareceres para um possível Sistema de Apoio à Decisão (SAD) utiliza-se *Data Warehouse* que podemos definir como um arcabouço temporal de dados, ou seja, acumulados durante um intervalo de tempo definido.

Kimball *et al.* [10] afirmam que um ambiente para ser definido como *Data Warehouse* - (DW) deve atender aos seguintes requisitos dos usuários:

- O sistema de DW/BI deve tornar a informação acessível, disponível e compreensível;
- Os dados devem ser intuitivos e de maneira óbvia para os usuários e possíveis tomadores de decisão de uma organização ou instituição;
- Estruturas e rótulos semelhantes a formas de pensar dos usuários de negócios, tanto em vocabulário quanto em formato simplificado e de fácil visualização;
- Possibilidade dos usuários de negócios separar e combinar dados analíticos com infinitas combinações;
- Desempenho em realização de consultas; e
- As ferramentas de Inteligência de Negócios - BI e aplicativos que acessam os dados devem ser simples e fácil de usar.

O *Data Warehouse* é o alicerce do processamento de dados analíticos. No ambiente analítico os dados destinados para a análise são obtidos a partir de uma fonte única, tornando a tarefa para análise SAD mais fácil e precisa que no ambiente operacional, onde as diversas fontes são independentes e não refletem uma realidade única [8].

### 3.3.1 Introdução à Modelagem Dimensional

A modelagem é uma técnica de projeto lógico que busca apresentar dados em uma estrutura padronizada de maneira intuitiva e com alto desempenho de acesso, de maneira a dar suporte para um processamento analítico [16]. Sendo a modelagem dimensional como uma alternativa para a modelagem entidade relacionamento com as mesmas informações. Tem-se o modelo dimensional composto por dois tipos de tabela [16] [10]:

**Fato** é a tabela central, possui uma chave primária composta que detem uma grande quantidade de dados - da ordem de *gigabytes* a *terabytes* - e possui as medições numéricas do negócio como o armazenamento das medições de desempenho resultantes a partir de eventos de processos de negócios de uma organização. Sendo que a chave primária simples da tabela dimensão está relacionada com a chave estrangeira da tabela fato gerando ligação entre elas.

A Figura 3.4 mostra a presença no centro do esquema estrela a tabela Fato, bem como cada linha da Fato corresponde a um evento de medição de dados, tendo que em que cada coluna possui um nível de especificidade de detalhe denominado grão, sendo um fundamento da modelagem dimensional de que todas as linhas de medições de uma Tabela Fato devem possuir mesmo grão.

**Dimensão** são tabelas que alimentam a tabela fato e comumente possuem uma chave simples, armazenam pequenas quantidades de dados comparado a tabela fato e o armazenamento é de dados descritivos do negócio. São componentes da tabela que relatam traços de uma coisa tangível, com padronizações de atributos textuais ao nomear as colunas que na tabela Fato podem ser substituídos por identificadores. As tabelas dimensões não variam com o tempo a cada nova inserção de novas linhas.

Ciente que a chave primária simples da tabela dimensão corresponde exatamente à chave estrangeira da tabela fato, permitindo, destarte a ligação entre as tabelas. No modelo dimensional todas as dimensões são análogas e possuem iguais pontos de entrada e inserção na tabela fato, conservando a simetria tanto para interfaces de usuário, estratégias de consultas e as próprias consultas SQL [10].

Kimball *et al.* [10] citam que a modelagem dimensional é amplamente aceita como técnica preferida para apresentar dados analíticos com a visão de apoiar um Sistema de Decisão, sendo dois requisitos simultâneos:

- Entrega de dados que sejam compreensível para usuários de negócios;
- Entrega de desempenho de consultas rápidas.

A modelagem dimensional possui como vantagem a versatilidade para acomodar nova inserção de dados, além de novas decisões de projetos [16], avante a capacidade de modificar as tabelas dimensões ao adicionar novas linhas de dados na tabela, não requerendo reprogramação das ferramentas de consultas e geradoras de relatórios analíticos para a introdução de novas modificações [10].

### 3.3.2 Sistemas OLTP e OLAP

Em geral, assumi-se que os sistemas OLTP fornecem dados de origem para *Data Warehouse*, no tempo em que os sistemas OLAP auxiliam na sua análise. Contudo, se destacam nas seguintes distinções:

- **Desempenho:** sistemas OLTP apresentam situação imediata dos negócios de uma organização, sendo atualizados continuamente à medida que as transações são realizadas [10], isto é, devem estar disponíveis no momento do acesso. Não obstante, sistemas OLAP realizam consultas sobre grandes massas de dados históricos de um evento ou situação, na qual deve apresentar a resposta no menor tempo possível para geração de relatórios analíticos, podendo esse tempo de processamento ser em segundos, minutos ou horas [10].
- **Acesso aos dados:** o DW é um repositório centralizado em que obtém e integra os mais diversos dados dos mais diversos bancos de dados de sistemas OLTP, assim como de fontes externas [10] [21]. O acesso aos relatórios analíticos deve ocorrer independentemente de quais sistemas OLTP as possuam [21].
- **Formato dos dados:** sistemas OLTP não possuem suporte explícito para representar corretamente um histórico de eventos passados e possuem dados nos mais diversos formatos.

Conquanto um sistema OLAP tanto quanto um DW possuem temporalidade (os dados representam uma granularidade de dia, mês ou ano) e sumarização dos dados para formar um banco de dados analíticos para geração de relatórios [10].

- **Qualidade dos dados:** em sistemas OLTP ocorrem atualizações constantes de status de um negócio à medida que as transações são concretizadas, o que em um DW tornaria inviável uma consulta analítica, pois de tempos em tempos vão mudando os dados e conseqüentemente as informações, produzindo resultados consideravelmente diferentes, não fornecendo dados generalistas.

O sistema OLAP fornece em tempo hábil de maneira mais concreta uma versão da realidade em um relatório analítico [21].

Em resumo as distinções se mostram mais claras com a tabela esquematizada por Otte *et al.* [17]:

OLTP	OLAP
Orientados a aplicações	Orientados a assuntos
As Vezes de Grande tamanho	Quase sempre grandes
Dados granulados	Dados constituídos de sumarizações
Dados de pouca fontes	Dados de múltiplas fontes
Suporta consultas e atualizações	Atualizações em modo <i>batch</i>
Dados que mudam constantemente	Dados mais estáveis
Dados atuais	Dados históricos

Figura 3.3: Distinções entre sistemas OLTP e OLAP. Conforme: Poe *et al.* [16].

d

### 3.3.3 Esquema Estrela comparado aos Cubos OLAP

Modelos dimensionais implementados em SGBD são referidos como esquemas em formatos de estrela por causa de sua grande semelhança com uma estrela - onde todas as tabelas estão desnormalizadas, tornando uma estrutura extremamente previsível oferece grande vantagem na geração de relatórios analíticos - sendo que o motor de banco de dados restringe inicialmente as tabelas dimensões para então realizar ações sobre a tabela fato uma única vez, na qual serão realizados as consultas [10].

O esquema Estrela ao ser hospedado em um banco de dados relacional é uma boa base física para a construção de um Cubo OLAP, que é geralmente considerado uma base mais estável para *backup* e recuperação [10].

Poe [16] elucida que a utilização de um esquema estrela Figura 3.4 para processamento analítico possibilita alguns benefícios em relação a uma estrutura relacional, dos quais destacam-se os aspectos a seguir:

- O projeto de banco de dados multidimensional provê rápido tempo de resposta;
- Permite otimizar o banco de dados para trabalhar com um projeto mais simples de banco de dados melhorando a execução do planejamento;
- Permite projetar o banco de dados como o usuário final habitualmente pensa e usa os dados analiticamente;
- Simplifica o entendimento e a navegação dos metadados por desenvolvedores e usuários finais;

- Possibilita um maior número de ferramentas de acesso aos dados, sendo que alguns produtos disponíveis no mercado exigem o projeto de um esquema estrela.

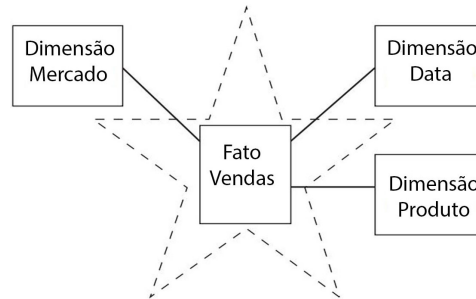


Figura 3.4: Esquema estrela de *Data Warehouse* - conforme Kimball *et al.* [10].

Modelos implementados em ambientes de bancos de dados multidimensionais são referidos cubos de processamento analítico on-line - OLAP. Quando carregados em um cubo OLAP, tem-se que o armazenamento de dados ocorre de maneira indexada usando formatos e técnicas que são concebidos para dados multidimensionais. Esse formato de indexação do Cubo fornece agregações de desempenho e tabelas de resumo pré-calculados gerenciáveis pelo próprio Cubo OLAP, em consequência, tem maior ganho de desempenho comparado ao modelo Estrela [10].

Os usuários podem usar o processo de *Driv Down* (que consiste em uma informação sumarizada ou indexada, diminuindo a granularidade de seus dados, podem ser visualizados informações mais detalhadamente, como de ano para dias) de modo a adicionar ou remover atributos de suas análise com maior desempenho de emissão de novas consultas. Cubos Olap fornecem consultas mais robustas quando comparado ao esquema Estrela e as disponíveis em uma consulta SQL. Entretanto, tais recursos que envolvam a granularidade de dados com processos de *Drill Down* em um Cubo Olap paga-se o preço em desempenho dado a quantidade de memória e processamento para geração de tais consultas analíticas, principalmente quando envoltos em uma grande quantidade de dados [10].

Mas conforme Kimball *et al.* [10] as tecnologias OLAP estão melhorando continuamente em aspectos de processamento, para acompanhar as necessidades que as informações sejam minuciosamente detalhadas, e que sejam carregadas em sua totalidade de maneira atômica no esquema Estrela, como na Figura 3.4.

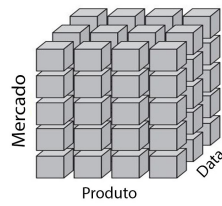


Figura 3.5: Cubo OLAP de *Data Warehouse* - conforme [10].

### 3.3.4 Processo de Extração, Transformação e Carga - ETL

A área de preparação de dados do *Data Warehouse* é tanto uma área de armazenamento quanto um conjunto de processos comumente referido como o extração-transformação-carga ETL, conforme [10]:

1. Os dados da área de preparação (transformação) são tudo aquilo que estão entre os sistemas e a fonte de dados operacionais da área de apresentação de dados;
2. No armazém de dados, dados operacionais bruto são transformados em um ajuste do armazém a ser entregue para consulta do usuário e do consumo.

# Capítulo 4

## Estudo de Caso

### 4.1 Do Objeto de estudo

### 4.2 Dados obtidos sobre a Educação Superior Brasileira

Os dados utilizados nesta pesquisa foram obtidos junto ao portal INEP [6], inicialmente, foram estudados os dados verificando as diferenças anos após ano, sendo os dados disponíveis para *Download* referente aos anos de 1996 até o ano de 2014, até a presente momento da publicação deste Projeto de Conclusão de Curso, ainda não foi disponibilizados o ano de 2016, após carga dos dados em *Data Warehouse* (DW) não estavam disponibilizado os dados do ano de 2015 no referido portal.

Etapa que consiste no processo de Extração, Transformação e Carga de Dados em um DW - a qual consistiu na etapa que mais demandou tempo - sendo uma etapa complexa e extremamente detalhada pela leitura da documentação e tentativa de compreensão de um modelo relacional para posteriormente torná-lo um modelo dimensional que viabiliza a carga em um DW. Mesmo com as várias ferramentas que possam vir a auxiliar ainda é um processo de grande necessidade de tempo e estudo, conforme Oliveira [15].

Sendo posteriormente verificado no processo de extração desses dados, forte variação dos levantamentos realizados de um período governamental para o outro, havendo assim intervenção política no tratamento desses dados do Ensino Superior Brasileiro.



## 4.3 Histórico da Instituição - INEP

### 4.3.1 Instituto Nacional de Estudos e Pesquisa Anísio Teixeira (INEP)

A instituição INEP, conforme o próprio sítio institucional [7], foi criado, por lei, no dia 13 de janeiro de 1937, sendo num primeiro momento chamado de Instituto Nacional de Pedagogia. Posteriormente um Decreto-Lei, instituiu regulamentação, organização e estrutura institucional, modificando sua denominação para Instituto Nacional De Estudos Pedagógicos.

Segundo esse Decreto-Lei [7], cabia ao INEP:

1. Organizar a documentação relativa à história e ao estado atual das doutrinas e técnicas pedagógicas;
2. Manter intercâmbio com instituições do País e fora dele;
3. Promover inquéritos e pesquisas;
4. Prestar assistência técnica aos serviços estaduais, municipais e particulares de educação, ministrando-lhes, mediante consulta ou independentemente dela, esclarecimentos e soluções sobre problemas pedagógicos;
5. Divulgar os seus trabalhos.

O INEP tinha intuito, após tentativas frustradas de sistematizar os conhecimentos educacionais e propositura de melhorias no ensino que já haviam sido tentadas, de ter a devida continuidade de estudos e pesquisas educacionais. Após esses fatores o Instituto Nacional de Estudos e Pesquisa Anísio Teixeira (INEP) se tornou referência no país.

Em 1952 [7], assumiu a direção da autarquia o professor Anísio Teixeira que era responsável pelos registros históricos educacionais do ensino brasileiro em todas suas dimensões. Adiante deu nome ao Inep que até o término deste projeto ainda é chamado Instituto Nacional de Estudos e Pesquisa Anísio Teixeira.

Em 1972, o INEP se torna órgão autônomo. Em 1985 tornar-se função de suporte e assessoramento aos processos decisórios do MEC, no período do governo Collor, quando quase foi extinto o referido instituto, após isso houve outro processo de reestruturação da missão, da autarquia, em dois objetivos [7]:

- Reorientação das políticas de apoio a pesquisas educacionais, buscando melhorar sua performance no cumprimento das funções de suporte à tomada de decisões em políticas educacionais;
- Reforço do processo de disseminação de informações educacionais, incorporando novas estratégias de modalidades de produção e difusão de conhecimentos e informações.

Entretanto, somente no ano de 1995, após uma intensa reestruturação, que o INEP começou atuar como financiador de trabalhos acadêmicos voltados para educação, que gerou grandes estudos e levantamentos estatístico, também sendo disponibilizados a partir desse ano estudo sobre a Educação Superior Brasileira, de tal modo que orientasse a Educação Brasileira como um todo na formulação de políticas públicas do MEC [7]. Atualmente o INEP é responsável pela maioria dos levantamentos estatístico referentes ao Ensino Superior - objeto dessa pesquisa - bem como de todo ensino brasileiro, incluindo o Enem, Prova Brasil e censo escolar.

### 4.3.2 Sobre o INEP

O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) é uma autarquia vinculada ao Ministério da Educação - MEC, que conforme o Portal do INEP [7], possui a seguinte missão:

- Promoção dos estudos, pesquisas e avaliações sobre o Sistema Educacional Brasileiro com o objetivo de subsidiar a formulação e implementação de políticas públicas para a área educacional a partir de parâmetros de qualidade e equidade, bem como produzir informações claras e confiáveis aos gestores, pesquisadores, educadores e público em geral.

Para gerar seus dados e estudos educacionais o INEP realiza levantamentos estatísticos e avaliativos em todos os níveis e modalidades de ensino, através de:

- **Censo Escolar:** levantamento de informações estatísticos-educacionais de âmbito nacional, levantamento realizado anualmente;
- **Censo Superior:** coleta, anualmente, uma série de dados do ensino superior no País, incluindo cursos de graduação, presenciais e à distância - **objeto de estudo deste trabalho de conclusão de curso.**

- **Avaliação dos Cursos de Graduação:** é um procedimento utilizado pelo MEC para o reconhecimento ou renovação de reconhecimento dos cursos de graduação representando uma medida necessária para a emissão de diplomas;
- **Avaliação Institucional:** compreende a análise dos dados e informações prestrados pelas Instituições de Ensino Superior (IES) no formulário eletrônico e a verificação, no próprio local, da realidade institucional, do seus cursos de graduação e de pós-graduação, de pesquisa e de extensão;
- **SINAES:** é o mais atual instrumento de avaliação superior do MEC. Formado por três componentes: avaliação de instituições, curso e desempenho dos estudantes;
- **ENEM:** exame de saída facultativo aos que concluíram e aos concluintes do ensino médio, que teve sua primeira aplicação no ano de 1997;
- **ENCCEJA:** É uma proposta de projeto do MEC de construir um padrão de referência de avaliação nacional para jovens e adultos que não puderam ter a conclusão de área de ensino na idade apropriada; e
- **SAEB:** pesquisa por amostragem, do ensino fundamental e médio, realizada a cada dois anos.

Além de promover encontros e palestras para discussão de temas educacionais e disponibilização de outras fontes de consultas sobre educação.

### 4.3.3 Processo de Downloads de Microdados do Ensino Superior Brasileiro

O processo de extração dos dados ocorreu mediante *Download* dos dados fornecidos no seguinte sítio da *Internet*: <http://portal.inep.gov.br/basica-levantamentos-acessar>

No portal do INEP estavam presentes os dados dos anos de 1995 a 2014 para baixar, conforme apresentados na Tabela 4.1, arquivos compactados no formato .zip.

Após baixados os arquivos, correspondente a parte de extração do processo ETL, foi realizado esquema do modelo relacional relacionando as chaves primárias e estrangeiras das tabelas, para posterior construção do Modelo Dimensional dos Dados dispostos da maneira mais detalhada possível.

Tabela 4.1: Tamanho dos arquivos compactados baixados.

<b>Dados</b>	<b>Ano de Referência</b>	<b>Tamanho em megabytes</b>
Mircrodados	1995	2.19 MB
Mircrodados	1996	2.32 MB
Mircrodados	1997	1.73 MB
Mircrodados	1998	6.72 MB
Mircrodados	1999	1.89 MB
Mircrodados	2000	4.52 MB
Mircrodados	2001	4.23 MB
Mircrodados	2002	5.99 MB
Mircrodados	2003	6.33 MB
Mircrodados	2004	7.74 MB
Mircrodados	2005	8.50 MB
Mircrodados	2006	11.7 MB
Mircrodados	2007	15.7 MB
Mircrodados	2008	14.1 MB
Mircrodados	2009	160.0 MB
Mircrodados	2010	356 MB
Mircrodados	2011	513 MB
Mircrodados	2012	805 MB
Mircrodados	2013	204 MB
Mircrodados	2014	259 MB

Em síntese conforme Kimball *et al.* [10] o poder analítico de um ambiente de DW/BI é diretamente proporcional à qualidade e profundidade da dimensão de seus atributos. Quanto mais tempo se investe fornecendo atributos de maneira mais detalhada em uma coluna de atributos, melhor será a possibilidade de extração de informações. Consequente essa robustez auxiliam no corte de um cubo OLAP.

# Capítulo 5

## Implementação do Estudo de Caso

### 5.1 Implementação do Estudo de Caso - Análise de Dados Abertos do Ensino Superior Brasileiro

Para o desenvolvimento de um DW, Kimball *et al.* propôs o seguinte ciclo, consoante Figura 5.1 abaixo:

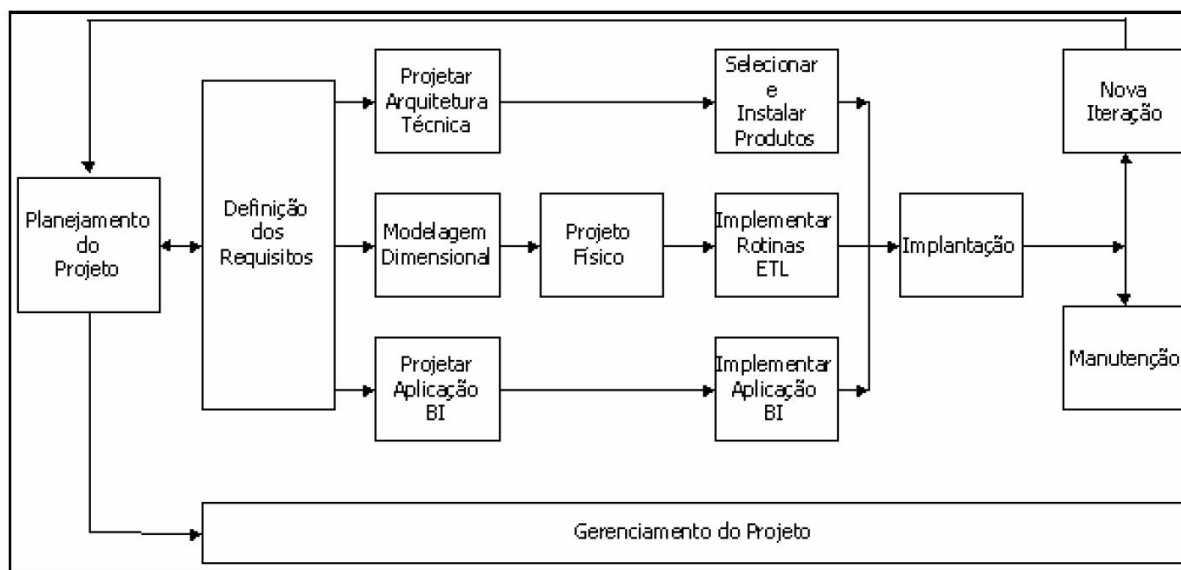


Figura 5.1: Ciclo de vida de um projeto de *Data Warehouse*, conforme: Kimbal *et al.* [10].

Os seguintes tópicos apresentam as etapas de elaboração do projeto de Análise de Dados Abertos do Ensino Superior Brasileiro adaptado ao ciclo de desenvolvimento de um DW proposto por Kimball *et al.* [10].

**Planejamento do Projeto:** atividade para elaboração do plano de projeto. Neste trabalho consistiu em selecionar a base de dados a ser iniciada o projeto e a definição do objeto a ser trabalhado no projeto e o planejamento para sua execução - na qual foi Análise de Dados Abertos com a obtenção de dados do Ensino Superior Brasileiro mostrando o detalhamento da execução das tarefas a serem realizadas na implementação desse Trabalho de Conclusão de Curso.

**Definição dos Requisitos:** trata-se do levantamento das necessidades do negócio e das fontes de dados. Ocorreu o processo de Levantamento das Necessidades do Negócio (a seleção para realizar esse trabalho foi a Análise de Dados Abertos do Ensino Superior) e o Levantamento das Fontes de Dados (Dados extraídos do portal INEP - [6]).

**Projeto de Arquitetura Técnica:** consiste na etapa de realizar o levantamento dos volumes envolvidos, na obtenção dos dados e seu respectivo uso de armazenamento, conforme Tabela 4.1 do Capítulo 4 bem como os *softwares* que irão compor as camadas de arquitetura OLAP.

**Modelagem Dimensional:** é a atividade de elaboração do modelo dimensional de dados em um *Data Warehouse* em acordo com os processos a serem executados e repetidos. Em conformidade com Kimball *et al.* [10] um modelo dimensional possui as mesmas informações quanto um modelo normalizado - o que proporciona maior compreensibilidade, consulta, desempenho e resistência a mudanças.

No presente projeto representa a realização de identificar as dimensões e fatos, bem como em cada fato são organizados as dimensões envolvidas, o conteúdo e a granularidade dos dados. Conforme Figura 5.2 a seguir:

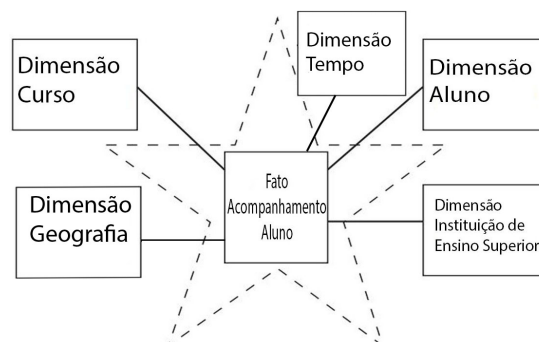


Figura 5.2: Esquema estrela de *Data Warehouse* utilizado na modelagem para Análise de Dados Abertos do Ensino Superior Brasileiro.

**Projeto de Aplicação BI:** prática que reside na definição de consultas previstas para iteração e a documentação dos indicadores a serem identificados. Processo que inclui levantamento das consultas e análises demandadas pelos usuários que serão a base para a aprovação do correspondente ciclo.

Neste projeto definiram-se as seguintes consultas e análises das possíveis demandas pelos usuários a serem obtidos da Tabela Fato, tais quais:

- Quantidade de alunos no ensino superior no decorrer dos anos;
- Sexo dos alunos do ensino superior por grupo de faixa etária;
- Quantidade de alunos matriculados por determinada grande área de conhecimento;
- Quantidade de alunos matriculados no Ensino Superior por curso;
- Níveis de titulação de docentes;
- Raça/cor dos alunos do Ensino Superior por Ano de Ingresso;
- Análise anual dos alunos do Ensino Superior formados por dada região;
- Situação acadêmica dos alunos no Ensino Superior por região.

**Projeto Físico:** consiste na construção do modelo físico relacional e multidimensional. O modelo físico deriva do modelo lógico e de tratamentos de desempenho e controle; o modelo multidimensional depende das visões e consultas a serem liberadas para os usuários - como consultas solicitadas, perfil de usuário e outros.

**Selecionar e Instalar Produtos:** constitui em preparar o ambiente de desenvolvimento, com a disponibilidade de *software* e *hardware* a serem configurados. As ferramentas utilizadas nesse trabalho são todos *softwares livres* sob a Licença Pública Geral - GPL, sendo elas:

1. **R:** para a transformação dos dados de .TXT para .CSV, que no início deste projeto estavam disponibilizados somente arquivos compactados que após descompactados estavam em .TXT;
2. **Pentaho Data-Inegration:** para a construção das rotinas de Extração, Transformação e Carga -ETL;
3. **Pentaho BI-Server:** para realização dos painéis de indicadores - *Dashboards*;
4. **MYSQL:** Sistemas Gerenciadores de de Banco de Dados (SGBD) para a criação de tabelas.

**Implementar Rotinas de ETL:** consiste na especificação e documentação dos processos de ETL dos dados. Inclui obtenção dos dados que alimentarão o DW a partir

da extração dos dados das bases transacionais, neste projeto foram obtidos pelos portal INEP via download [6].

**Implementar Aplicação de BI:** etapa que implementa consultas e das fórmulas que gerarão os indicadores solicitados pelos usuário. Neste trabalho foi implantado consultas e *Dashboards* (painel de indicadores) identificados e solicitadas pelos usuários.

**Implantação:** consiste na disponibilização de temas desenvolvidos na iteração para o DW em ambiente de produção e avaliação de desempenho. Esta fase inclui também a passagem da aplicação do ambiente de desenvolvimento para o produção, a documentação do processo e o treinamento dos profissionais, técnicos e especialistas que irão suprir à aplicação. Etapa que não foi realizada neste trabalho.

**Nova Iteração:** consiste no planejamento da próxima iteração considerando as necessidades dos usuários da aplicação e desenvolvedores após iterações já finalizadas. Sendo este o processo de retroalimentação do sistema, observando as necessidades vindouras - que no projeto de Análise de Dados Abertos sobre o Ensino Superior Brasileiro seria a alimentação de dados de anos posteriores e adequação da inserção de novos dados.

**Manutenção:** a manutenção irá ocorrer com a inserção de dados nas tabelas dimensões, com suas devidas correções e aceites para alimentar a tabela Fato em anos vindouros.

**Gerenciamento do Projeto:** concentra as tarefas relacionadas ao acompanhamento e ao controle da execução do projeto.

### 5.1.1 Implementação do processo de Extração, Transformação e Carga - ETL

Após definir o escopo do projeto como a obtenção de relatórios analíticos (*Dashboards*) a partir de dados abertos do Ensino Superior Brasileiro em um *Data Warehouse*, foi realizado o processo de *download* junto ao portal do INEP [6] para implementação de rotinas de ETL, em consonância com o proposto na imagem abaixo para construção dos relatórios:



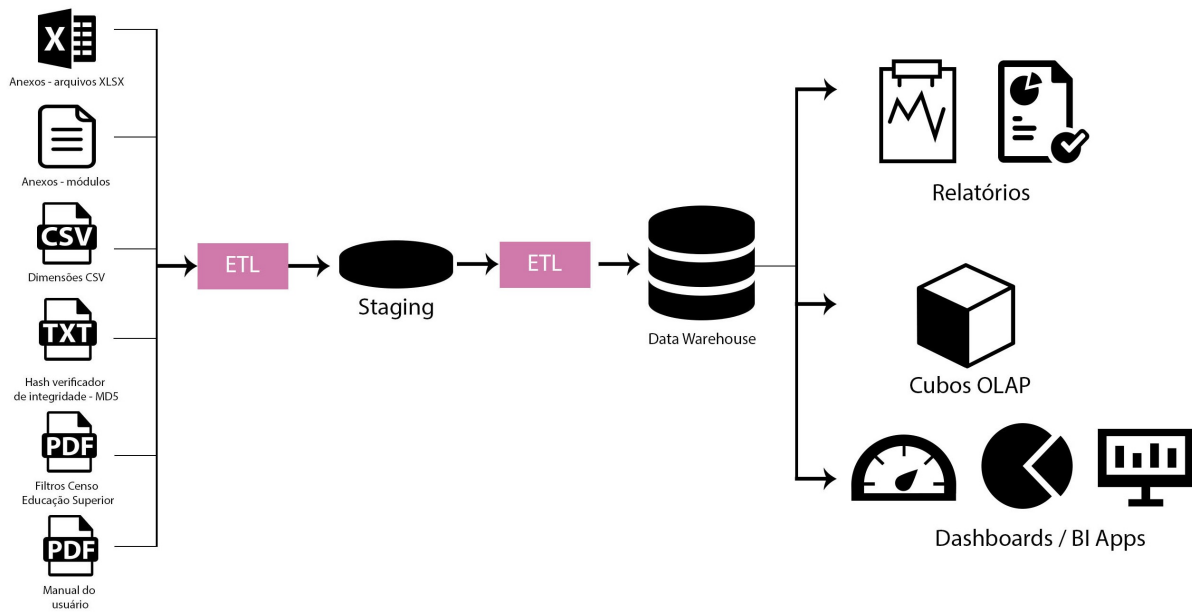


Figura 5.3: Modelo implementado em acordo com a definição do escopo do projeto para geração de relatórios analíticos.

### 5.1.2 Processo de Extração:

A extração é o primeiro passo no processo de obtenção de dados para o ambiente de *Data Warehouse*. Sendo a base escolhida para a realização deste estudo o Censo de Educação do Ensino Superior Brasileiro dentre os dados fornecidos pelo INEP, tendo como produto a obtenção de análises gerais do dados extraídos e realizar a carga dos dados necessários para um armazém de dados (como um banco de dados convencional, neste trabalho foi utilizado o SGBD *MYSQL*) para um ambiente de teste para posterior manipulação e obtenção de um possível SAD [10].

O objetivo do trabalho consiste em mostrar que é possível a geração de relatórios analíticos a partir de fonte de dados abertos, principalmente os governamentais, com o objetivo de construir *Dashboards* (painel de indicadores), para serem utilizados como uma ferramenta de apoio a decisão - Sistema de Apoio à Decisão (SAD) mostrados no Capítulo 3. Neste processo de extração do projeto, verificou os seguintes detalhes:

1. Os dados disponibilizados no próprio portal do INEP se encontravam a princípio compactados no formato *.ZIP* que ao descompactá-los estavam no formato *.TXT* referentes ao período de 1995 a 2013.

2. Separados esses dados ano a ano, conforme pode se verificar no portal [6], como visto na Tabela 4.1, contudo ocorreu atualização da página do portal [6] sendo disponibilizados os anos de 2014 e 2015, que estes já se encontravam no formato .CSV.
3. Utilizou-se *software* R para realizar uma programação a fim de transformar os dados em formato .TXT para .CSV.

Essas transformações são precursoras para carga de dados na área de apresentação de um *Data Warehouse*.

## 5.2 Processo de Transformação:

Uma vez que os dados são extraídos para a área de organização, existem inúmeros potenciais de transformações, através da realização da limpeza dos dados, conforme Kimball *et al.* [10] - tendo etapas como:

- Correção de erros de ortografia - neste projeto na extração dos dados do Inep [6] encontrados erros, tais como, de acentuação e uso de padrão UTF-8, que tem a função de representar qualquer caracter universal (aceitando a língua portuguesa) de padrão do formato Unicode, sendo então compatível com o ASCII adequando aos caracteres dos atributos das entidades (tabelas) referentes à aluno, curso, instituição de ensino superior, acompanhamento de aluno e de docente;
- Resolução de conflitos de domínio;
- Elementos em falta de modo a dar coesão aos dados;
- Analisar os formatos para assumirem um padrão - combinando dados de várias fontes;
- Duplicação de dados, que ao serem analisados em nossos estudos possuem diversas redundâncias para melhorar o desempenho de acesso a banco de dados no modelo extraído do INEP, todavia para a carga dos dados em uma Tabela Fato não se fazem necessário tais redundâncias e nem mesmo etapas de normalização;
- Atribuição de chaves primárias na referida Tabela Fato.

### **5.3 Cargas realizadas na STAGE para serem carregadas nas tabelas dimensões**

Processo realizado tendo como ano de referência os Microdados Censo Superior 2014, dado sua completude comparado com os outros anos, para a devida carga na tabela Dimensão e ser alimentado na construção da tabela fato, a partir do modelo conceitual de DW estrela - Star Schema. O código SQL encontra-se em Apêndice neste trabalho, com a construção das seguintes dimensões:

- **Tabela Dimensão Aluno**
- **Tabela Dimensão Curso**
- **Tabela Dimensão Geografia**
- **Tabela Dimensão Instituição de Ensino Superior**
- **Tabela Fato Acompanhamento de Aluno**
- **Tabela Dimensão Acompanhamento de Docentes**

# Capítulo 6

## Resultados

### 6.1 Resultados Obtidos após a realização de consultas

Como a equipe possuía prévio conhecimento dos dados após verificação dos Módulos de extração de dados do Censo da Educação Superior, que acompanhava os arquivos compactados extraídos do portal INEP, para a realização do projeto de Análise de Dados Abertos sobre o Ensino Superior Brasileiro, teve como objetivo a propositura de perguntas estratégicas a serem respondidas ao final do projeto para o devido entendimento do escopo do projeto na implementação do *Data Warehouse*.

Consoante o que foi visto na etapa do **Projeto de Aplicação BI** no Capítulo 5 em que foram definidas as perguntas estratégicas para realização de consultas para um Sistema de Apoio à Decisão conforme imagens dos *Dashboards* realizadas no *Software Pentaho* para geração dos seguintes relatórios:

1. **Quantidade de alunos formados por região;**

A obtenção desse *Dashboard* mostra o quantitativo de alunos formados no Ensino Superior Brasileiro. Percebe-se que o quantitativo de alunos aumentou significativamente durante os anos, bem como também existe uma grande concentração em determinadas regiões, tais como sudeste, centro-oeste e sul.

Sendo os anos de 2009 e 2010 os mais significativos no quantitativo de alunos formados, possivelmente alunos formados no ensino técnico, devido a investimentos Governamentais como o PRONATEC.

Cursos que normalmente possuem duração de 2 anos a 2 anos e meio, com formação rápida e área específica de atuação, provavelmente, uma das proposta de governo para inserção do jovem o mais rápido possível no mercado de trabalho. Configurando-se grande possibilidade de confronto de dados da formação do ensino superior com dados de imposto de renda ou dados de empregados contratados para geração de relatórios mais versáteis ainda. Figura 6.1

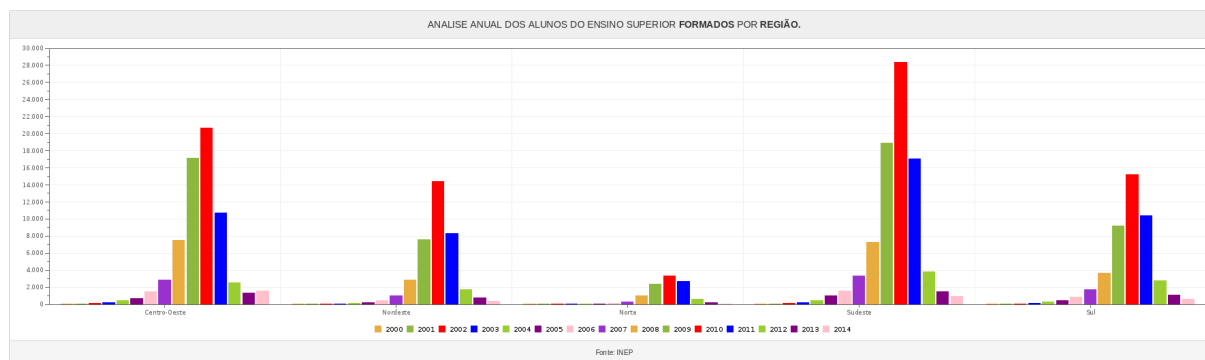


Figura 6.1: Quantitativo de alunos matriculados no ensino superior por ano de ingresso

## 2. Quantidade de alunos matriculados no Ensino Superior por ano

Verifica-se o aumento vertiginoso no quantitativo de alunos no Ensino Superior Brasileiro após o ano de 2008, provavelmente, devido a investimentos em educação pelo MEC quanto pelo próprio governo federal, mediante programas de financiamento estudantil como o FIES. Figura 6.2

Até o término deste trabalho não foram plenamente disponibilizados os dados de 2015 e os dados de 2016 encontram-se em aberto, não tendo o quantitativo exato de alunos.

Podendo confrontar o quantitativo de alunos com bases de dados referente aos repasses para educação, mensurando o custo dos alunos.

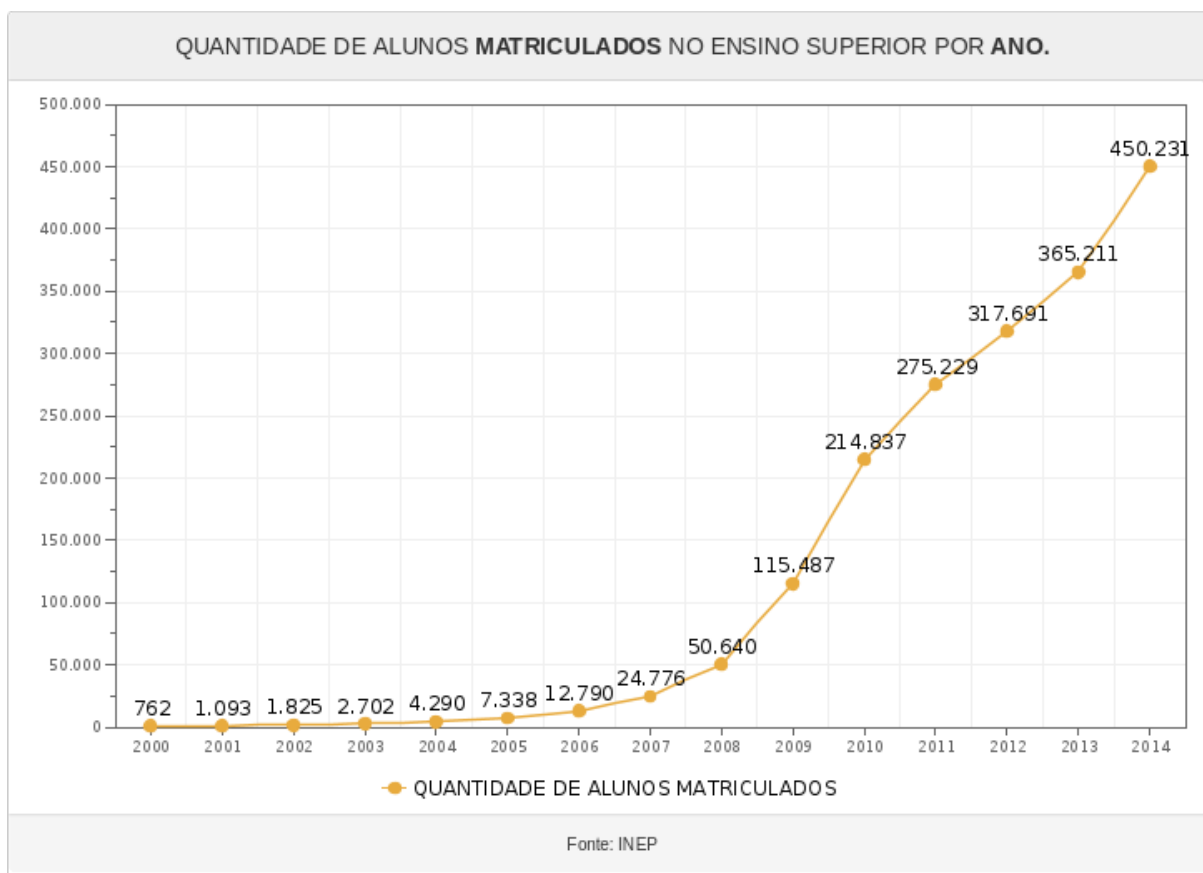


Figura 6.2: Quantitativo de alunos matriculados no ensino superior por ano de ingresso

### 3. Quantidade de alunos matriculados por área de conhecimento

Demonstrativo de *Dashboard* comparativo de quantitativo de alunos matriculados por grande área de conhecimento Figura 6.3, nas áreas de:

Agricultura e veterinária;  
 Ciências sociais, negócio e direito;  
 Ciências, matemática e computação;  
 Educação;  
 Engenharia, produção e construção;  
 Humanidades e artes;  
 Saúde e bem estar social;  
 Serviços.

Verifica-se que a quantidade de alunos se encontram-se em maior quantidade na área de Ciências Sociais, negócio e direito, na qual se encontram os cursos de direito e administração, que são os cursos com maiores quantidades de alunos. Figura 6.3

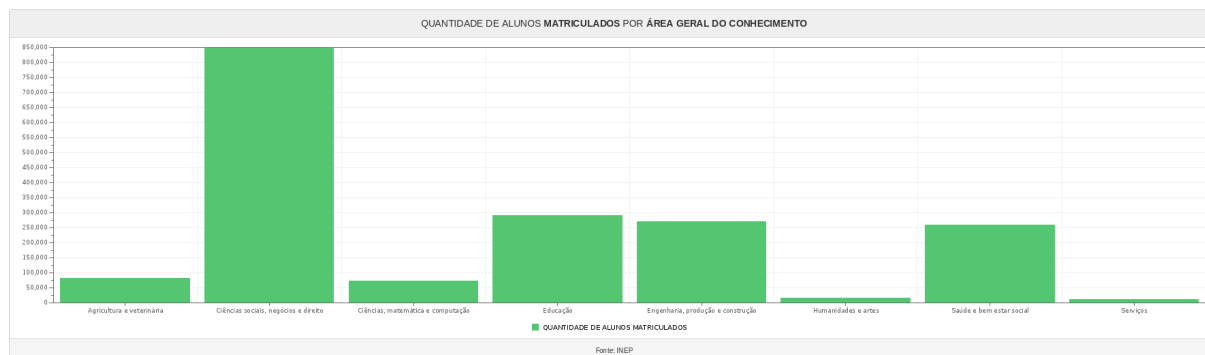


Figura 6.3: Quantitativo de alunos matriculados por grande área de conhecimento

4. Quantidade de alunos matriculados no Ensino Superior por curso;

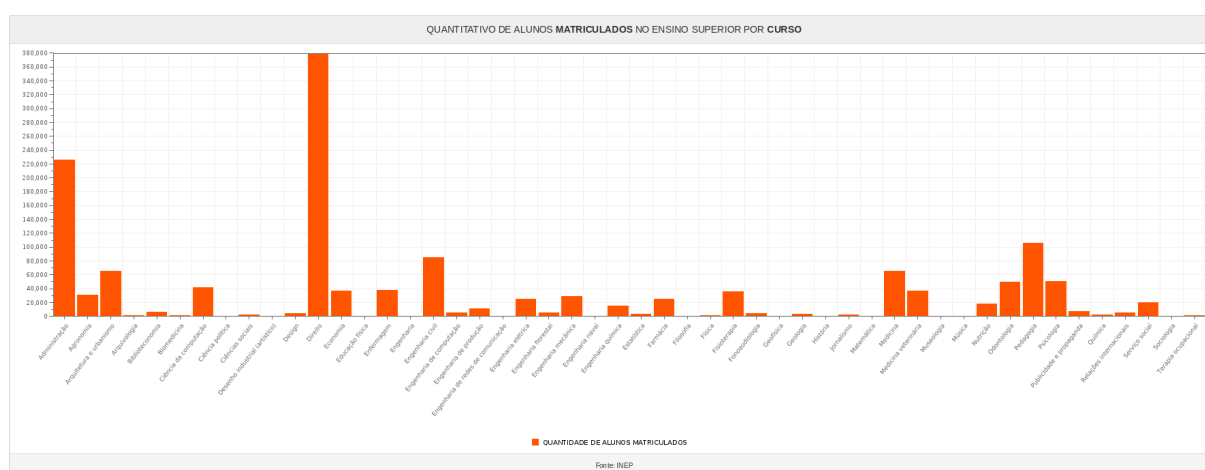


Figura 6.4: Quantitativo de alunos matriculados no ensino superior por ano de ingresso

Demonstrativo mostrando os cursos com o maior quantitativo de alunos por curso, sendo os 5 cursos com maiores quantidades de alunos são os cursos de:

- (a) Direito;
- (b) Administração;
- (c) Pedagogia;
- (d) Engenharia Civil;
- (e) Medicina.

## 5. Raça/cor dos alunos do Ensino Superior por Ano de Ingresso;

O critério raça-cor realizado de acordo com os critérios definidos pelo IBGE que são: Amarela, Branca, Indígena, Parda e Preta; por fim um critério caso o aluno não queira declarar cor/raça. Figura 6.5

Percebe-se a gradatividade do aumento da quantidade de negros e pardos nas instituições de ensino superior, possivelmente, pelos programas de cotas por raça/cor; entretando prevalece a quantidade de brancos e alunos que não quiseram declarar sua raça/cor; conquanto que a raça/cor indígena se manteve constante ao longo dos anos de 2010 e 2014, mas foi crescendo de maneira homogênea durante os outros anos.

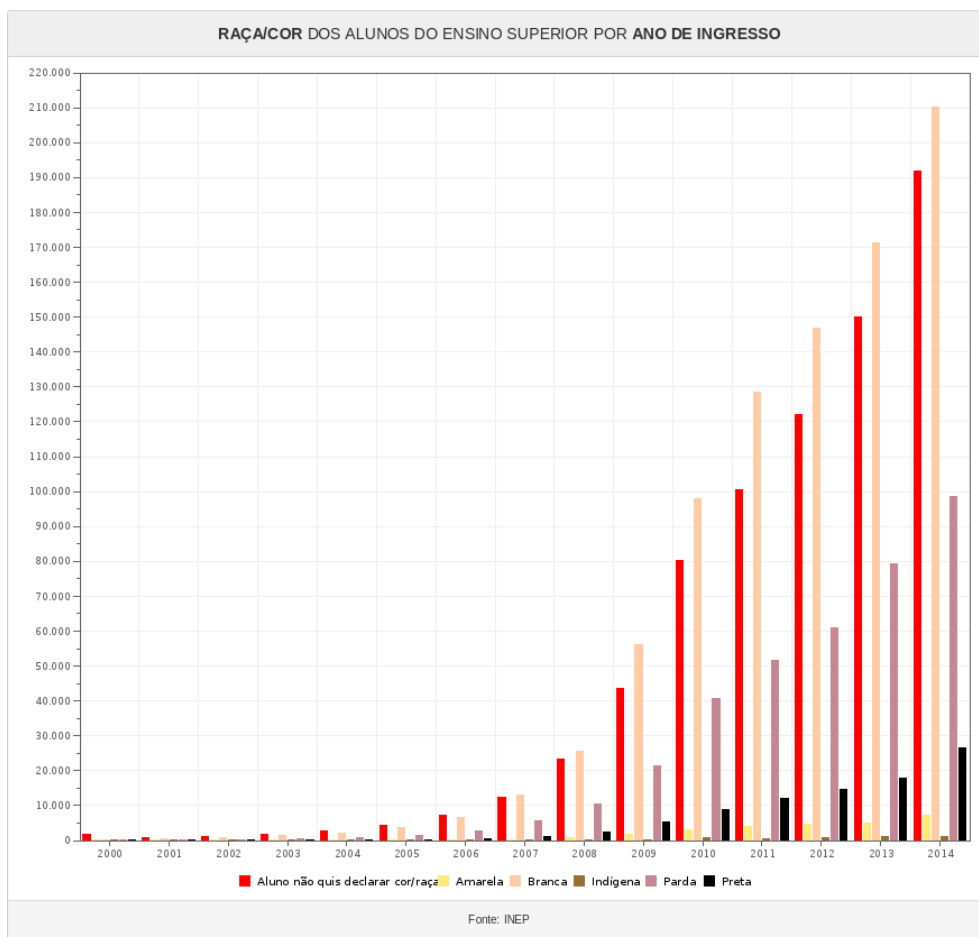


Figura 6.5: Quantitativo de alunos matriculados no ensino superior por raça/cor

## 6. Alunos matriculados por faixa etária e sexo;

Verifica-se que a quantidade de alunos do gênero feminino prevalece no ensino superior brasileiro, de tal maneira que a quantidade de mulheres aos 21 anos de



idade são mais que o dobro da quantidade de homens nas Instituições de Ensino Superior.

Também é verificado que o ingresso no ensino superior ocorre em maior quantidade na faixa etária de 18 aos 25 anos de idade, representando mais de 75% de toda a amostra analisada. Figura 6.6

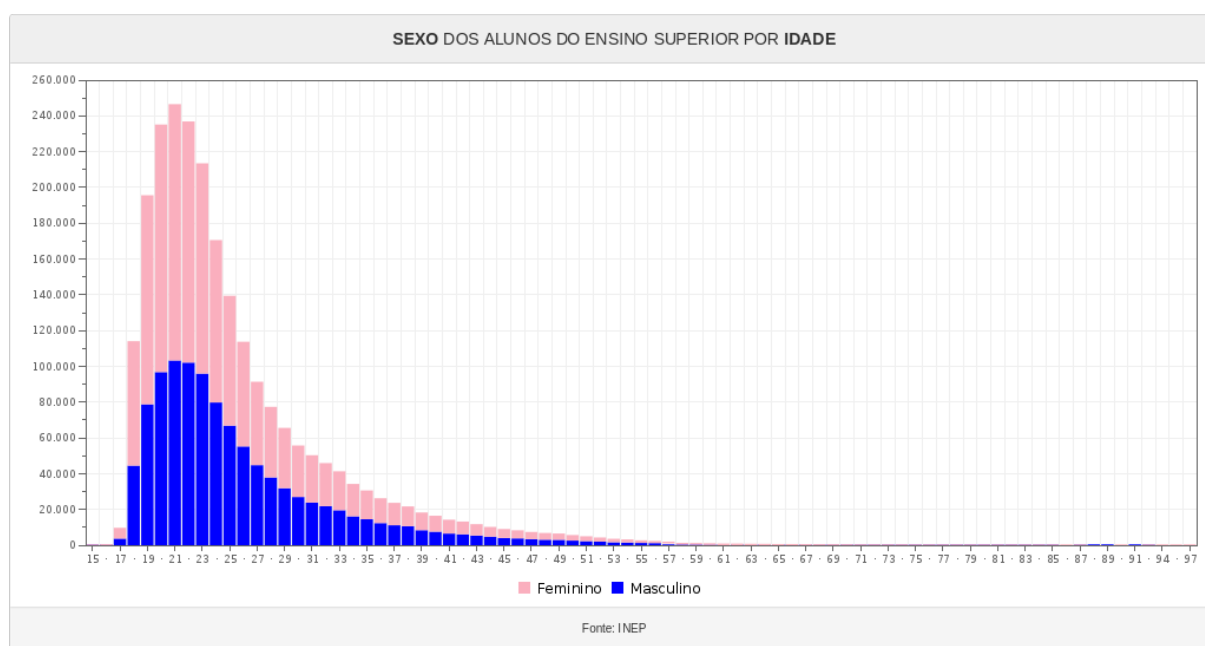


Figura 6.6: Quantitativo de alunos matriculados no ensino superior por ano de ingresso por sexo e faixa etária

## 7. Quantitativo de alunos por sexo e Unidade da Federação

Afere que tanto por estados quanto na dimensão global de todo ensino superior brasileira ocorre o fato que tem mais pessoas do gênero feminino no Ensino Superior do que do gênero masculino, sendo o único estado a ter praticamente igualado este quantitativo é o estado de Sergipe Figura 6.7.

Também apurou que o estado que possui maior quantidade de alunos no ensino superior é o estado de São Paulo, em seguida o Rio de Janeiro e Minas Gerais, ao mesmo tempo que o estado de Roraima possui o menor quantitativo de alunos no ensino superior.

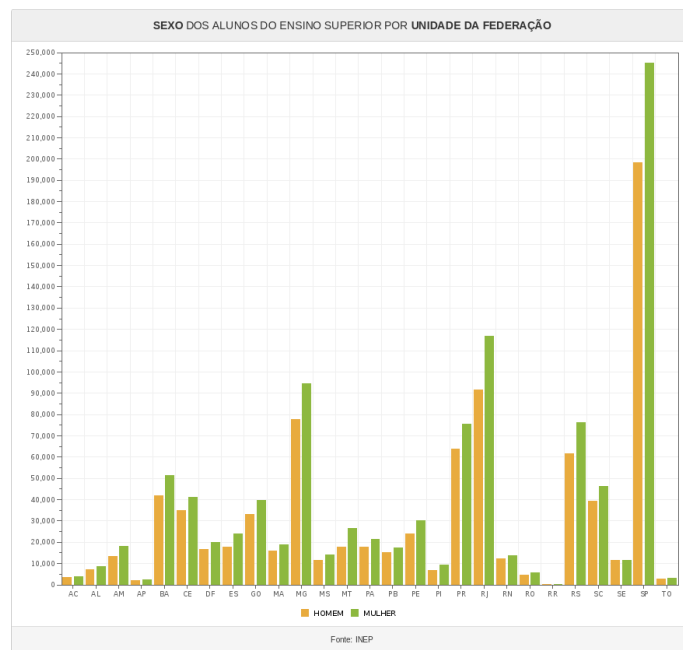


Figura 6.7: Quantitativo de alunos por sexo e estado

## 8. Situação acadêmica dos alunos no Ensino Superior por região.

As respectivas regiões da situação acadêmica dos alunos são: sul, suldeste, norte, nordeste e centro-oeste. Com a definição da seguinte situação acadêmica dos alunos:

- a - Cursando;
- b - Desvinculado do curso;
- c - Falecido;
- d - Formado;
- e - Matrícula trancada;
- f - Transferido para outro curso da mesma Instituições de Ensino Superior.

Na qual constata-se que as regiões que possuem maior de desvinculação com o curso são as regiões Sudeste e Centro-oeste, entretanto são as regiões com maior quantidade de alunos cursando o ensino superior.

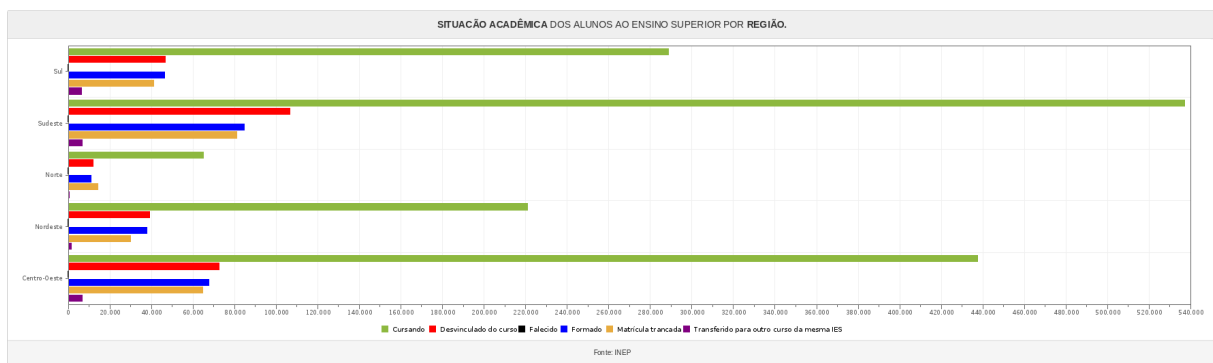


Figura 6.8: Situação Acadêmica dos alunos por região

# Capítulo 7

## Conclusões e Trabalhos Futuros

### 7.1 Considerações Finais

Não se pode duvidar que a divulgação de dados governamentais de forma aberta incrementa a transparência da administração pública e pode proporcionar inúmeros benefícios aos governos e cidadãos. Todavia, não é uma tarefa simples processar o volume massivo e crescente dos dados gerados pelo governo brasileiro, a fim de gerar *insights* úteis a gestores públicos, cidadãos e organizações interessadas.

Neste íterim, o presente trabalho apresentou uma abordagem integrada para a criação de um Sistema de Apoio à Decisão para a análise de dados abertos governamentais utilizando ferramentas livres.

A aplicação desenvolvida tem por objetivo disponibilizar relatórios analíticos a partir do processamento de dados abertos governamentais integrados em um DW.

Pretende-se, com o presente trabalho proporcionar ao gestor público realizar análises em busca de *insights* que possam ajudar no estabelecimento de políticas públicas eficazes, proporcionando o emprego racional de recursos públicos para o desenvolvimento do País; além de incentivar a população em geral no acompanhamento das políticas públicas por meio do acesso irrestrito, no ambiente *Web*, a dados abertos governamentais consistentes.

### 7.2 Trabalhos Futuros

Para dar continuidade a este trabalho, são sugeridos os seguintes estudos:

- Incluir novos dados no DW criado oriundos de outras áreas governamentais a fim de realizar cruzamentos de dados de diversas áreas proporcionando *insights* mais ricos.
- Criar políticas de publicação de dados abertos governamentais para facilitar o processo de ETL em uma arquitetura OLAP.
- Utilizar o banco de dados *NoSQL* para o armazenamento dos dados do DW para comparar o desempenho com banco de dados *MySQL*.
- Criar uma aplicação de mineração de dados, a partir dos dados do DW, a fim de empregar técnicas de aprendizado computacional para analisar e extrair automaticamente conhecimento dos dados existentes.

# Referências

- [1] World Wide Web Consortium et al. **Best practices for publishing linked data**. 2014. 6
- [2] Daniel Dietrich, Jonathan Gray, Tim McNamara, Antti Poikola, P Pollock, Julian Tait, e Ton Zijlstra. **Open data handbook**, 2009. 5
- [3] David Eaves. **The three laws of open government data**. *Eaves. ca*, 30, 2009. 8, 9
- [4] Stephen R Gardner. **Building the Data Warehouse**. *Communications of the ACM*, 41(9):52–60, 1998. 15
- [5] Dados Gov. **O que são dados abertos?** <http://dados.gov.br/dados-abertos>, 2013. 1, 5, 9
- [6] INEP. **Microdados para download**. <http://portal.inep.gov.br/basicalevantamentos-acessar>, 2015. 2, 22, 28, 30, 32
- [7] Portal INEP. **Conheça o INEP**. <http://portal.inep.gov.br/conheca-o-inep>, 2011. 1, 23, 24
- [8] William H Inmon, Ken Rudin, Christopher K Buss, e Ryan Sousa. **Data warehouse performance**. Wiley Publishing, 1998. 15, 16
- [9] Seiji Isotani e Ig Ibert Bittencourt. **Dados Abertos Conectados**. Novatec Editora, 2015. 6, 7
- [10] Ralph Kimball e Margy Ross. **The data warehouse toolkit: the complete guide to dimensional modeling**. John Wiley e Sons, 2011. 1, 15, 16, 17, 18, 19, 20, 21, 26, 27, 28, 31, 32
- [11] Lawrence Zordam Klein, Maria Luiza Machado Campos, Astério Kiyoshi Tanaka, e Praça General Tibúrcio. **A Tecnologia Objeto-Relacional em Ambientes de Data Warehouse: Uso de Séries Temporais como Tipo de Dado Não Convencional**. *Anais do XIV Simpósio Brasileiro de Banco de Dados*, pages 365–378, 1999. 15
- [12] Kenneth C Laudon e Jane P Laudon. **Sistemas de informação gerenciais**. São Paulo, 5ª Edição, Pearson, 2004. 5, 10, 11, 12, 13, 14

- [13] Felipe Nery Rodrigues Machado. *Projeto de banco de dados: uma visão prática*. Ed. Érica, 1996. 4
- [14] Ian Abramson Michael Corey, Michael Abbey e Ben Taub. *A Practical Guide to Successful Data Warehouse Analysis, Build and Roll Out*, 1998. 15
- [15] Wilson José de Oliveira. *Data warehouse*. Florianópolis/SC: Editora Visual Books, 2002. 22
- [16] Vidette Poe, Stephen Brobst, e Patricia Klauer. *Building a data warehouse for decision support*. Prentice-Hall, Inc., 1997. 14, 17, 18, 19
- [17] Paul Patrick Randy Otte e Mark Roy. *Understanding CORBA: Common Object Request Broker Architecture*. Prentice Hall PTR, 1996. 19
- [18] Peter Rob e E Coronel. *Sistemas de Banco de Dados-Projeto, Implementação e Administração*. 2011. 5
- [19] Arun Sen e Varghese S Jacob. **Industrial-strength data warehousing: why process is so important—and so often ignored**. *Communications of the ACM*, 41(9):29–31, 1998. 15
- [20] TCU. *5 Motivos para abertura de dados abertos na Administração Pública*, volume 1. 2015. 9
- [21] Hugh J Watson e Paul Gray. *Decision support in the data warehouse*. Prentice Hall Professional Technical Reference, 1997. 18, 19

## Apêndice A

*Dashboards*, criação de tabelas,  
diagrama ER, formulários de  
preenchimento



***Dashboards* em maior escala para melhor visualização.**

Segue a realização de *Dashboards* em modo paisagem para melhor visualização das informações sobre o Ensino Superior Brasileiro.

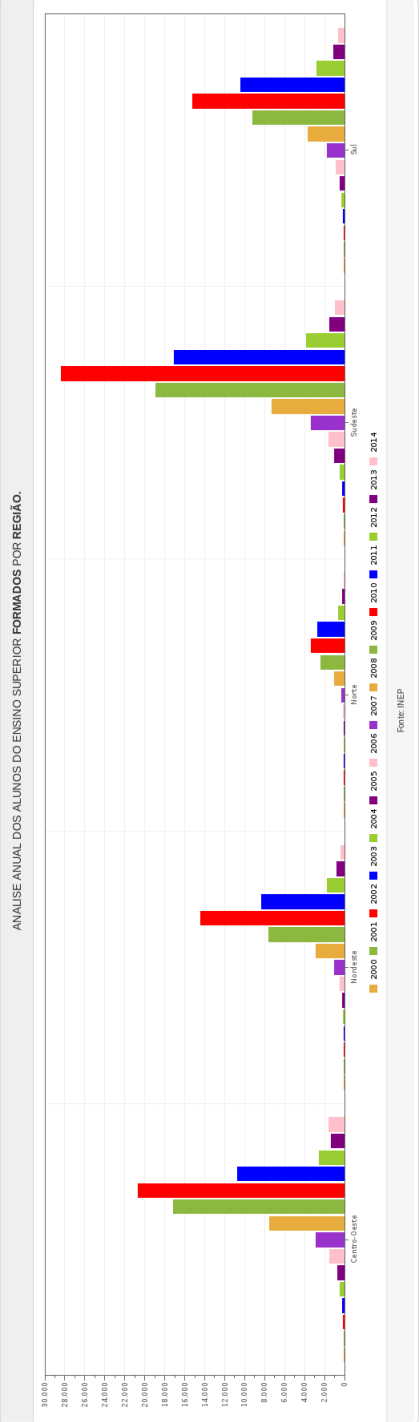


Figura 1: Análise anual dos alunos do ensino superior formados por região

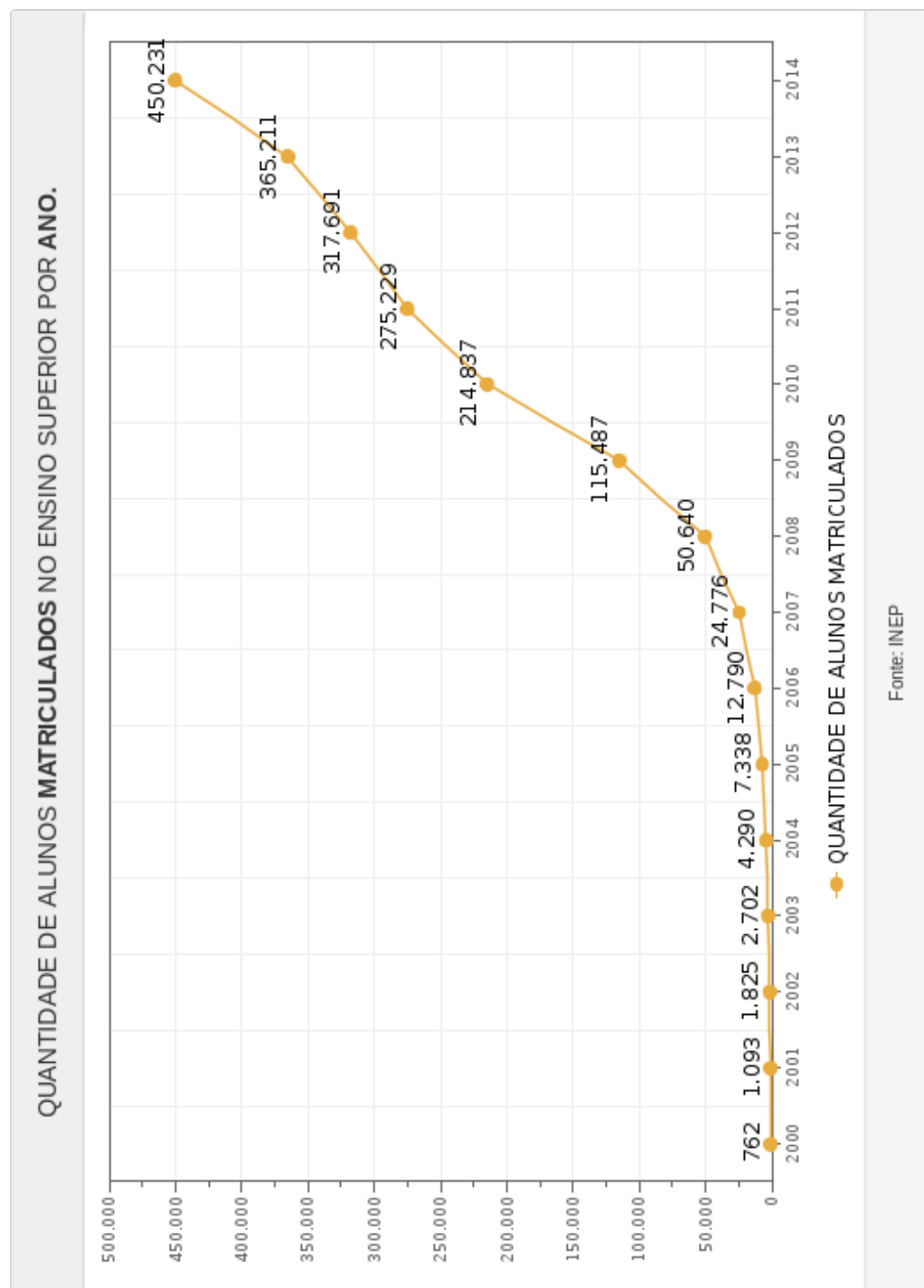


Figura 2: Quantidade de alunos matriculados no ensino superior por ano

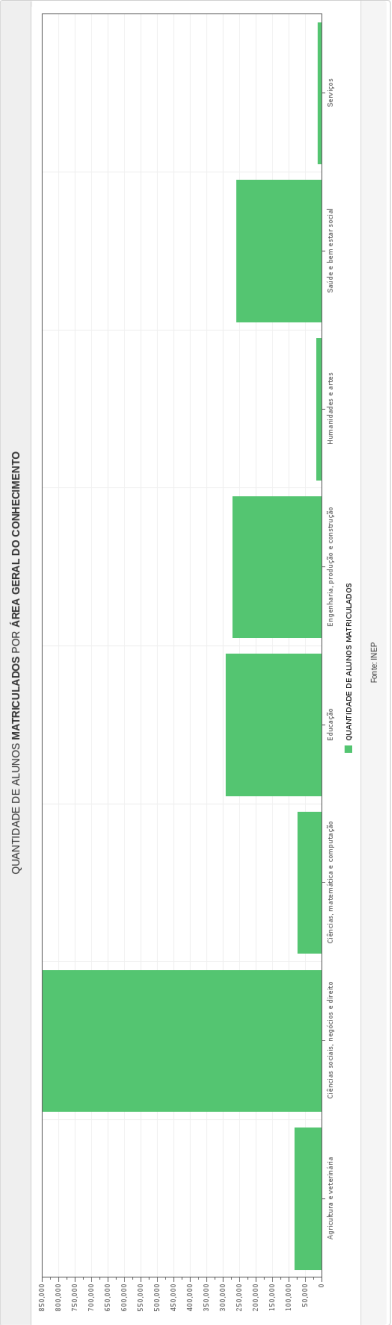


Figura 3: Quantitativo de alunos matriculados por área geral do conhecimento

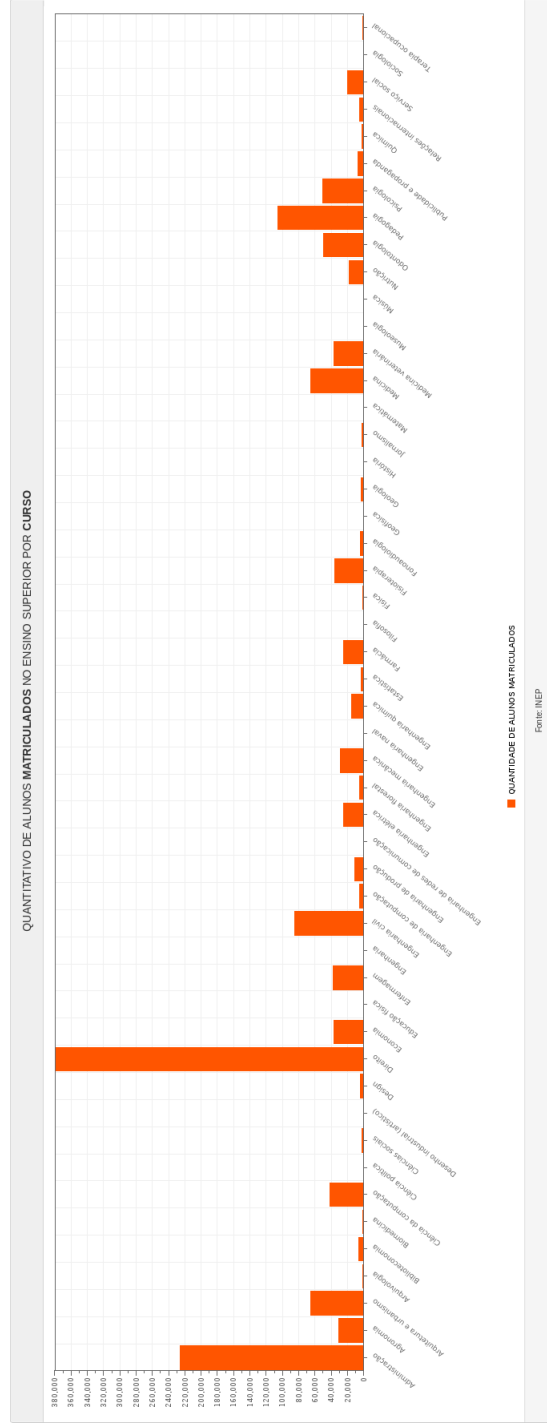


Figura 4: Quantitativo de alunos matriculados no ensino superior por curso

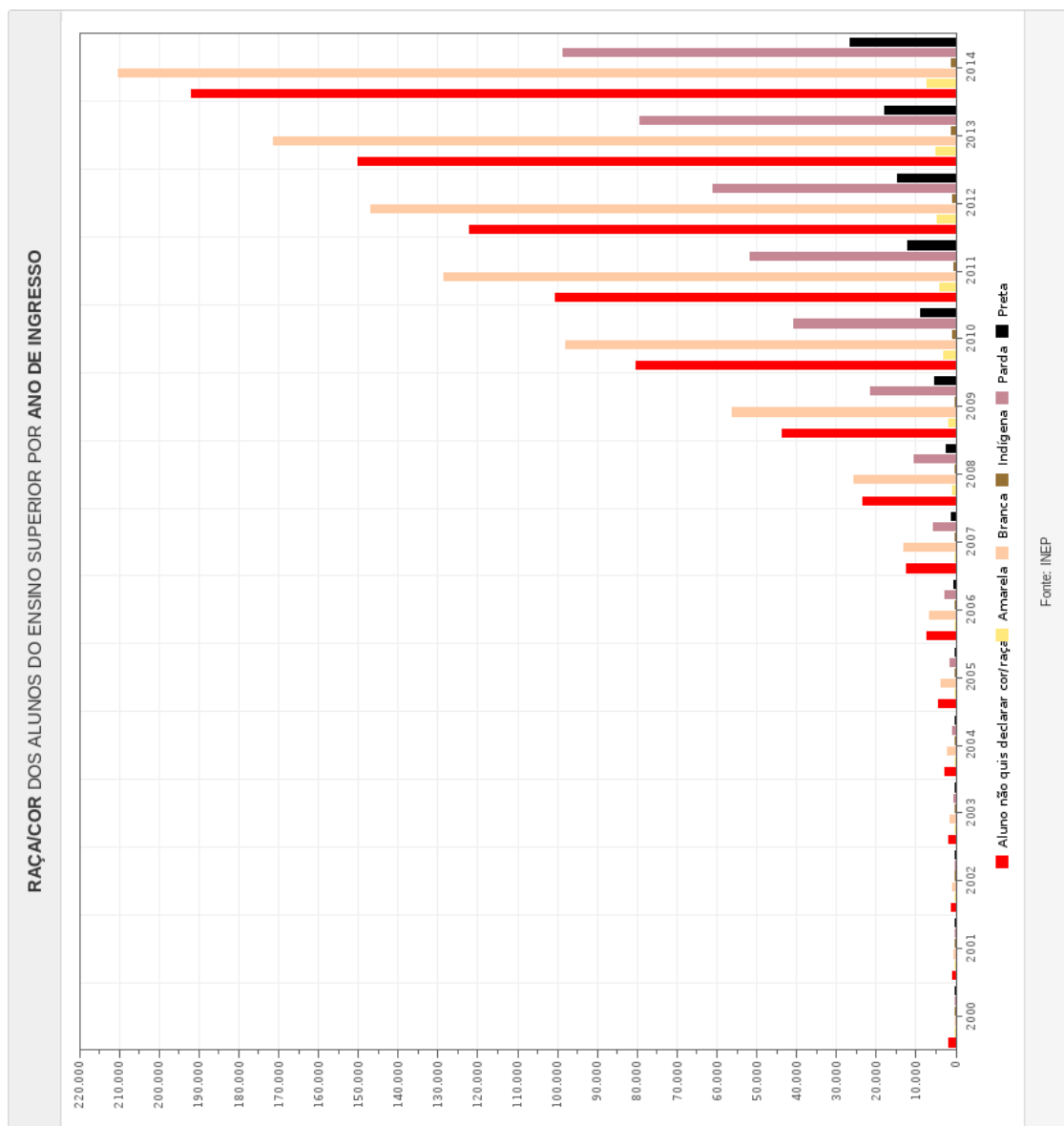


Figura 5: Raça/cor dos alunos matriculados no ensino superior

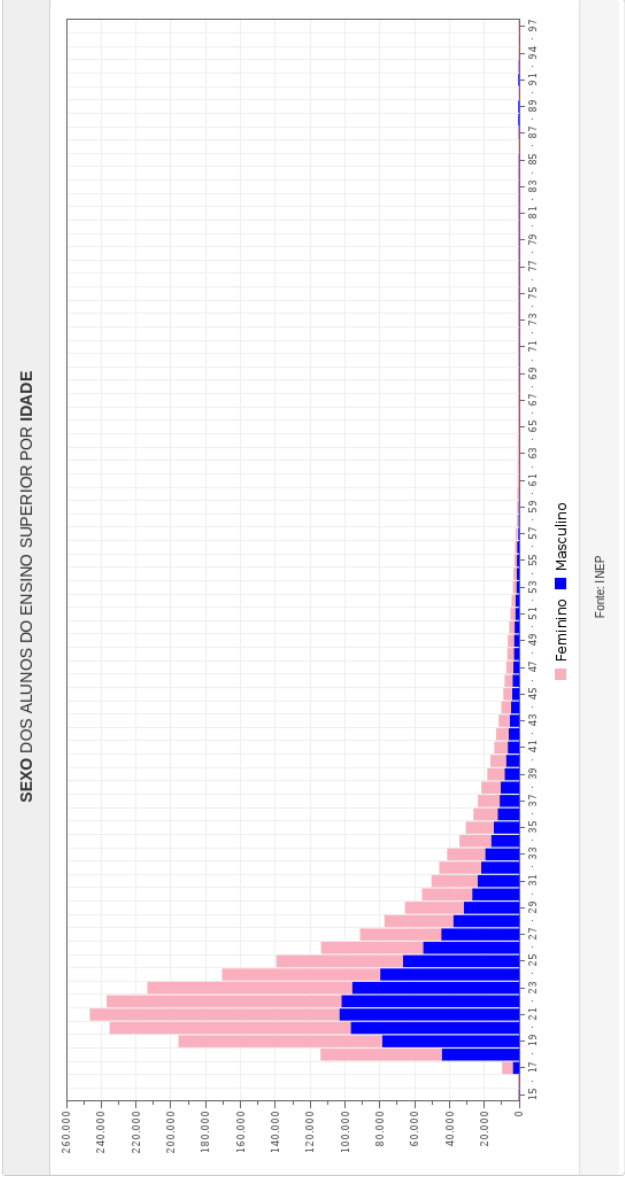


Figura 6: Sexo dos alunos matriculados no ensino superior por faixa etária

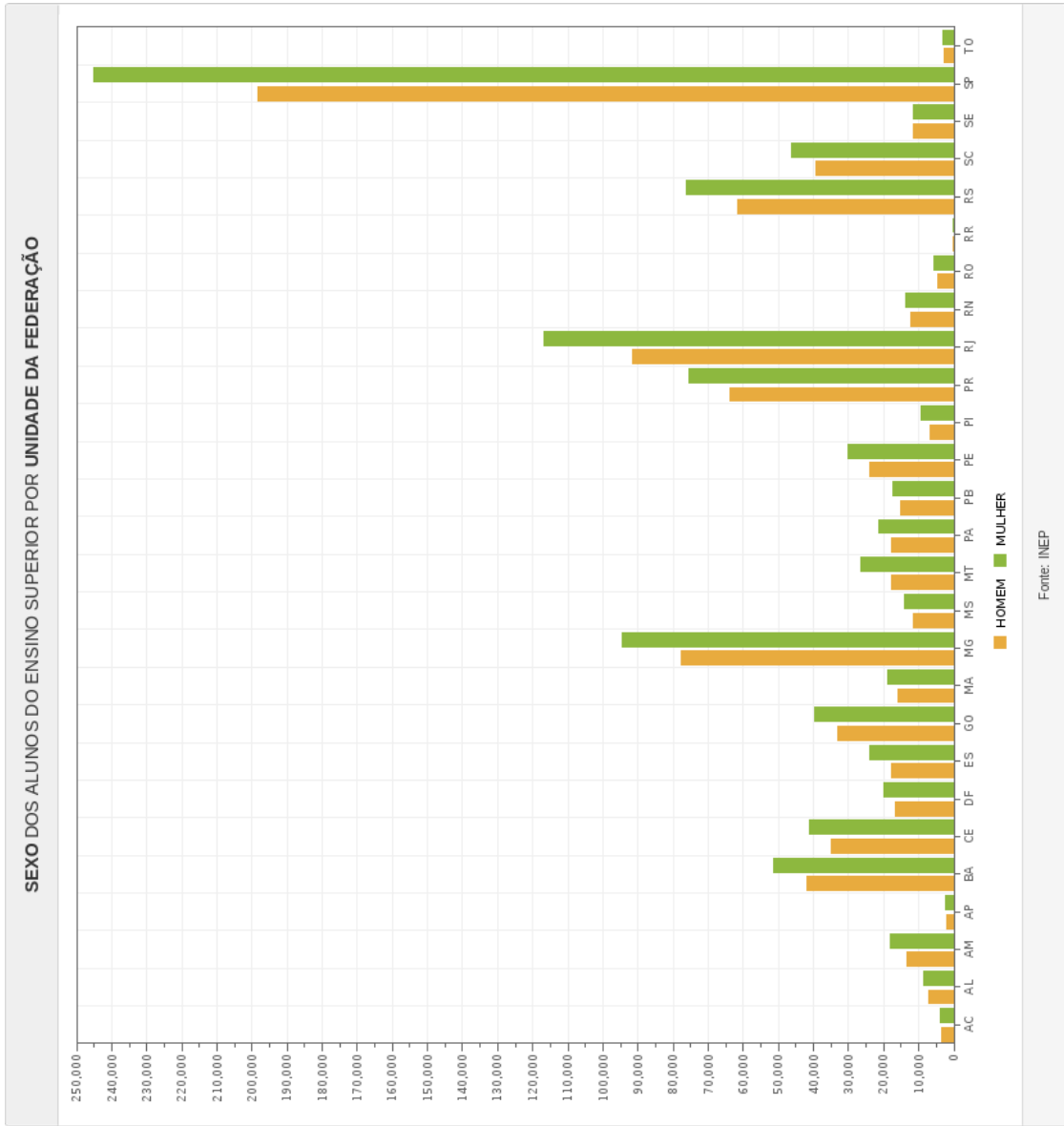


Figura 7: Sexo dos alunos matriculados no ensino superior por Unidade da Federação



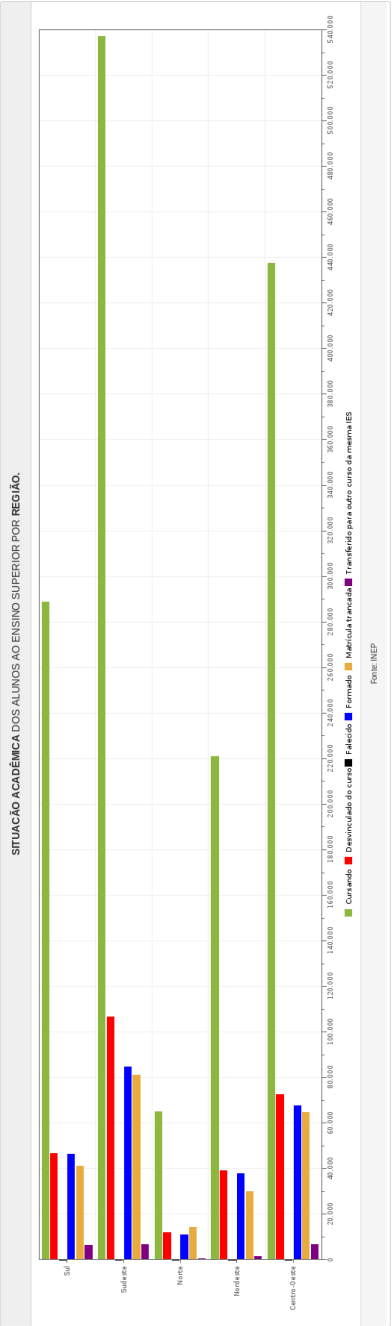


Figura 8: Situação dos alunos no ensino superior brasileiro

### Cargas realizadas na STAGE para serem carregadas nas tabelas dimensões

Processo realizado tendo como ano de referência os Microdados Censo Superior 2014, dado sua completude comparado com os outros anos, para a devida carga na tabela Dimensão e ser alimentado na construção da tabela fato, a partir do modelo conceitual de DW estrela - Star Schema. Salienta-se que o atributo ENGINE=InnoDB DEFAULT CHARSET=latin1 ao final do processo de criação das tabelas é para definir o caracteres especiais da língua portuguesa que estão contidas nos dados.

1. **Tabela Dimensão Aluno:** tabela criada no BD, com o seguinte comando SQL, para dimensão aluno 2014 e suas respectivas colunas com seus respectivos atributos:

```
CREATE TABLE DIM_ALUNO
(
ALUNO_COD_SEQ bigint PRIMARY KEY,
CO_ALUNO_CURSO bigint,
CO_ALUNO_CURSO_ORIGEM bigint,
CO_ALUNO bigint,
CO_COR_RACA_ALUNO bigint,
DS_COR_RACA_ALUNO VARCHAR(32),
IN_SEXO_ALUNO bigint,
DS_SEXO_ALUNO VARCHAR(9),
NU_ANO_ALUNO_NASC bigint,
NU_MES_ALUNO_NASC bigint,
NU_DIA_ALUNO_NASC bigint,
NU_IDADE_ALUNO bigint,
CO_NACIONALIDADE_ALUNO bigint,
DS_NACIONALIDADE_ALUNO VARCHAR(48),
CO_PAIS_ORIGEM_ALUNO bigint,
CO_UF_NASCIMENTO bigint,
CO_MUNICIPIO_NASCIMENTO bigint,
IN_ALUNO_DEF_TGD_SUPER bigint,
CO_ALUNO_SITUACAO bigint,
DS_ALUNO_SITUACAO VARCHAR(41),
DT_INGRESSO_CURSO VARCHAR(18),
IN_ING_VESTIBULAR bigint,
IN_ING_ENEM bigint,
IN_ING_AVALIACAO_SERIADA bigint,
IN_ING_SELECAO_SIMPLIFICADA bigint,
IN_ING_SELECAO_VAGA_REMANESC bigint,
IN_ING_SELECAO_VAGA_PROG_ESPEC bigint,
IN_ING_TRANSF_EXOFFICIO bigint,
IN_ING_DECISAO_JUDICIAL bigint,
IN_ING_CONVENIO_PECG bigint,
IN_RESERVA_VAGAS bigint,
IN_RESERVA_ETNICO bigint,
IN_RESERVA_DEFICIENCIA bigint,
IN_RESERVA_ENSINO_PUBLICO bigint,
```

```

IN_RESERVA_RENDA_FAMILIAR bigint,
IN_RESERVA_OUTRA bigint,
IN_FINANC_ESTUDANTIL bigint,
IN_APOIO_SOCIAL bigint,
IN_ATIVIDADE_EXTRACURRICULAR bigint,
CO_TIPO_ESCOLA_ENS_MEDIO bigint,
CO_SEMESTRE_CONCLUSAO bigint,
CO_SEMESTRE_REFERENCIA bigint,
IN_MOBILIDADE_ACADEMICA bigint,
CO_IES_DESTINO bigint,
CO_PAIS_DESTINO bigint,
IN_MATRICULA bigint,
IN_CONCLUINTE bigint,
IN_INGRESSO_TOTAL bigint,
IN_INGRESSO_VAGA_NOVA bigint,
ANO_INGRESSO bigint
)
ENGINE=InnoDB DEFAULT CHARSET=latin1;

```

2 **Tabela Dimensão Curso:** tabela criada no BD, com seguinte o comando SQL, para dimensão curso e suas respectivas colunas com seus respectivos atributos:

```

CREATE TABLE DIM_CURSO
(
CURSO_COD_SEQ bigint PRIMARY KEY,
CO_CURSO bigint,
NO_CURSO VARCHAR(158),
CO_SITUACAO_CURSO bigint,
DS_SITUACAO_CURSO VARCHAR(12),
CO_OCDE VARCHAR(6),
NO_OCDE VARCHAR(83),
CO_OCDE_AREA_GERAL bigint,
NO_OCDE_AREA_GERAL VARCHAR(36),
CO_OCDE_AREA_ESPECIFICA bigint,
NO_OCDE_AREA_ESPECIFICA VARCHAR(44),
CO_OCDE_AREA_DETALHADA bigint,
NO_OCDE_AREA_DETALHADA VARCHAR(55),
CO_GRAU_ACADEMICO bigint,
DS_GRAU_ACADEMICO VARCHAR(13),
CO_MODALIDADE_ENSINO bigint,
DS_MODALIDADE_ENSINO VARCHAR(17),
CO_NIVEL_ACADEMICO bigint,
DS_NIVEL_ACADEMICO VARCHAR(33),
IN_GRATUITO bigint,
TP_ATRIBUTO_INGRESSO bigint,
CO_LOCAL_OFERTA bigint,
NU_CARGA_HORARIA bigint,
DT_INICIO_FUNCIONAMENTO VARCHAR(18),
DT_AUTORIZACAO_CURSO VARCHAR(18),

```

```

IN_AJUDA_DEFICIENTE bigint,
IN_INTEGRAL_CURSO bigint,
IN_MATUTINO_CURSO bigint,
IN_VESPERTINO_CURSO bigint,
IN_NOTURNO_CURSO bigint,
NU_INTEGRALIZACAO_INTEGRAL bigint,
NU_INTEGRALIZACAO_MATUTINO bigint,
NU_INTEGRALIZACAO_VESPERTINO bigint,
NU_INTEGRALIZACAO_NOTURNO bigint,
NU_INTEGRALIZACAO_EAD bigint,
NU_PERC_CAR_HOR_SEMI_PRES bigint,
IN_POSSUI_LABORATORIO bigint,
QT_INSC_VAGAS_NOVAS_INT bigint,
QT_INSC_VAGAS_NOVAS_MAT bigint,
QT_INSC_VAGAS_NOVAS_VESP bigint,
QT_INSC_VAGAS_NOVAS_NOT bigint,
QT_INSC_VAGAS_NOVAS_EAD bigint,
QT_INSC_VAGAS_REMAN_INT bigint,
QT_INSC_VAGAS_REMAN_MAT bigint,
QT_INSC_VAGAS_REMAN_VESP bigint,
QT_INSC_VAGAS_REMAN_NOT bigint,
QT_INSC_VAGAS_REMAN_EAD bigint,
QT_INSC_VAGAS_PROG_ESP_INT bigint,
QT_INSC_VAGAS_PROG_ESP_MAT bigint,
QT_INSC_VAGAS_PROG_ESP_VESP bigint,
QT_INSC_VAGAS_PROG_ESP_NOT bigint,
QT_INSC_VAGAS_PROG_ESP_EAD bigint,
QT_VAGAS_NOVAS_INTEGRAL bigint,
QT_VAGAS_NOVAS_MATUTINO bigint,
QT_VAGAS_NOVAS_VESPERTINO bigint,
QT_VAGAS_NOVAS_NOTURNO bigint,
QT_VAGAS_NOVAS_EAD bigint,
QT_VAGAS_REMANESC_INTEGRAL bigint,
QT_VAGAS_REMANESC_MATUTINO bigint,
QT_VAGAS_REMANESC_VESPERTINO bigint,
QT_VAGAS_REMANESC_NOTURNO bigint,
QT_VAGAS_REMANESC_EAD bigint,
QT_VAGAS_PROG_ESP_INTEGRAL bigint,
QT_VAGAS_PROG_ESP_MATUTINO bigint,
QT_VAGAS_PROG_ESP_VESPERTINO bigint,
QT_VAGAS_PROG_ESP_NOTURNO bigint,
QT_VAGAS_PROG_ESP_EAD bigint,
QT_MATRICULA_CURSO bigint,
QT_CONCLUINTE_CURSO bigint,
QT_INGRESSO_CURSO bigint,
QT_INGRESSO_VAGAS_NOVAS bigint
)
ENGINE=InnoDB DEFAULT CHARSET=latin1;

```

3 **Tabela Dimensão Geografia:** tabela criada no BD, com o seguinte comando

SQL, para dimensão geografia e as respectivas colunas Município, Unidade da Federação, Região, País e seus respectivos atributos:

```
CREATE TABLE DIM_GEOGRAFIA
(
GEOGRAFIA_COD_SEQ bigint PRIMARY KEY,
COD_MUNICIPIO bigint,
NOME_MUNICIPIO VARCHAR(30),
COD_UF bigint,
SIGLA_UF VARCHAR(2),
NOME_REGIAO VARCHAR(12),
CAPITAL bigint,
COD_PAIS bigint,
NOME_PAIS VARCHAR(30)
)
ENGINE=InnoDB DEFAULT CHARSET=latin1;
```

4 **Tabela Dimensão Instituição de Ensino Superior:** tabela criada no BD, com o seguinte comando SQL, para dimensão referente ao Instituição de Ensino Superior e suas respectivas colunas com seus respectivos atributos:

```
CREATE TABLE DIM_IES
(
IES_COD_SEQ bigint PRIMARY KEY,
CO_IES bigint,
NO_IES VARCHAR(100),
SGL_IES VARCHAR(20),
CO_MANTENEDORA bigint,
CO_CATEGORIA_ADMINISTRATIVA bigint,
DS_CATEGORIA_ADMINISTRATIVA VARCHAR(27),
CO_ORGANIZACAO_ACADEMICA bigint,
DS_ORGANIZACAO_ACADEMICA VARCHAR(51),
CO_MUNICIPIO_IES bigint,
NO_MUNICIPIO_IES VARCHAR(27),
CO_UF_IES bigint,
SGL_UF_IES VARCHAR(2),
NO_REGIAO_IES VARCHAR(12),
IN_CAPITAL_IES bigint,
QT_TEC_TOTAL bigint,
QT_TEC_FUND_INCOMP_FEM bigint,
QT_TEC_FUND_INCOMP_MASC bigint,
QT_TEC_FUND_COMP_FEM bigint,
QT_TEC_FUND_COMP_MASC bigint,
QT_TEC_MEDIO_FEM bigint,
QT_TEC_MEDIO_MASC bigint,
QT_TEC_SUPERIOR_FEM bigint,
QT_TEC_SUPERIOR_MASC bigint,
QT_TEC_ESPECIALIZACAO_FEM bigint,
QT_TEC_ESPECIALIZACAO_MASC bigint,
QT_TEC_MESTRADO_FEM bigint,
```

```

QT_TEC_MESTRADO_MASC bigint,
QT_TEC_DOUTORADO_FEM bigint,
QT_TEC_DOUTORADO_MASC bigint,
IN_ACESSO_PORTAL_CAPES bigint,
IN_ACESSO_OUTRAS_BASES bigint,
IN_REFERENTE bigint,
VL_RECEITA_PROPRIA bigint,
VL_TRANSFERENCIA bigint,
VL_OUTRA_RECEITA bigint,
VL_DES_PESSOAL_REM_DOCENTE bigint,
VL_DES_PESSOAL_REM_TECNICO bigint,
VL_DES_PESSOAL_ENCARGO bigint,
VL_DES_CUSTEIO bigint,
VL_DES_INVESTIMENTO bigint,
VL_DES_PESQUISA bigint,
VL_DES_OUTRAS bigint
)
ENGINE=InnoDB DEFAULT CHARSET=latin1;

```

1. **Tabela Dimensão Tempo:** tabela criada no BD, com o seguinte comando SQL, para dimensão Tempo, com as colunas Ano e tempo, bem como seus respectivos atributos:

```

CREATE TABLE DIM_TEMPO
(
TEMPO_COD_SEQ bigint PRIMARY KEY,
ANO bigint,
SEMESTRE bigint
)
ENGINE=InnoDB DEFAULT CHARSET=latin1;

```

- 5 **Tabela Fato Acompanhamento de Aluno:** tabela criada no BD, com seguinte comando SQL, para Tabela Fato referente ao acompanhamento do aluno, de modo a verificar seu ano de ingresso no curso e na IES, além de responder as perguntas estratégicas definidas na etapa do Projeto de Aplicação de BI, bem como suas respectivas colunas com seus respectivos atributos:

```

CREATE TABLE FATO_ACOMPANHAMENTO_ALUNO
(
DIM_ALUNO_COD_SEQ bigint,
DIM_CURSO_COD_SEQ bigint,
DIM_IES_COD_SEQ bigint,
DIM_TEMPO_COD_SEQ bigint,
DIM_GEOGRAFIA_ALUNO_COD_SEQ bigint,
DIM_GEOGRAFIA_IES_COD_SEQ bigint,
PRIMARY KEY (DIM_ALUNO_COD_SEQ, DIM_CURSO_COD_SEQ,
DIM_IES_COD_SEQ, DIM_TEMPO_COD_SEQ,
DIM_GEOGRAFIA_ALUNO_COD_SEQ,
DIM_GEOGRAFIA_IES_COD_SEQ),

```

```

FOREIGN KEY (DIM_ALUNO_COD_SEQ) REFERENCES DIM_ALUNO (ALUNO_COD_SEQ),
FOREIGN KEY (DIM_CURSO_COD_SEQ) REFERENCES DIM_CURSO (CURSO_COD_SEQ),
FOREIGN KEY (DIM_IES_COD_SEQ) REFERENCES DIM_IES (IES_COD_SEQ),
FOREIGN KEY (DIM_TEMPO_COD_SEQ) REFERENCES DIM_TEMPO (TEMPO_COD_SEQ),
FOREIGN KEY (DIM_GEOGRAFIA_ALUNO_COD_SEQ) REFERENCES DIM_GEOGRAFIA (GEOGRAFIA_COD_SEQ),
FOREIGN KEY (DIM_GEOGRAFIA_IES_COD_SEQ) REFERENCES DIM_GEOGRAFIA (GEOGRAFIA_COD_SEQ)
)
ENGINE=InnoDB DEFAULT CHARSET=latin1;

```

**6 Tabela Dimensão Acompanhamento de Docentes:** tabela de criação no BD, com seguinte comando SQL, para dimensão referente ao acompanhamento dos docentes, de modo a verificar a lotação na respectiva IES, e respectivas ofertas de disciplinas e curso oferecidos pelos mesmos, além de suas respectivas colunas com seus respectivos atributos:

```

CREATE TABLE
    FT_ACOMPANHAMENTO_DOCENTE
(
CO_IES bigint,
CO_DOCENTE_IES bigint,
CO_DOCENTE bigint,
CO_LOCAL_OFERTA bigint
)
ENGINE=InnoDB DEFAULT CHARSET=latin1;

```

Não houve a carga de Dimensões na Tabela Fato de Acompanhamento de Docentes por ausência de máquinas para dar suporte a referida carga, enquanto eram trabalhados a carga das outras dimensões referentes a Tabela Fato de Acompanhamento do Aluno que demandou grande uso de memória volátil e processamento, tornando um inviabilizador para realizar a carga das dimensões que alimentariam a Tabela Fato Acompanhamento Docente como outro gerador de relatórios analíticos na implementação do referido projeto.

Modelo de consulta de seleção com técnica *Look up Table*, na qual cria-se uma tabela como novos valores ou inserção deles para geração de uma tabela somente de consultas.

```

SELECT
ALUNO.ALUNO_COD_SEQ AS DIM_ALUNO_COD_SEQ,
CURSO.CURSO_COD_SEQ AS DIM_CURSO_COD_SEQ,
IES.IES_COD_SEQ AS DIM_IES_COD_SEQ,
TEMPO.TEMPO_COD_SEQ AS DIM_TEMPO_COD_SEQ,
GEOALUNO.GEOGRAFIA_COD_SEQ AS DIM_GEOGRAFIA_ALUNO_COD_SEQ,
GEOIES.GEOGRAFIA_COD_SEQ AS DIM_GEOGRAFIA_IES_COD_SEQ

FROM STG_ALUNO AS ALUNO

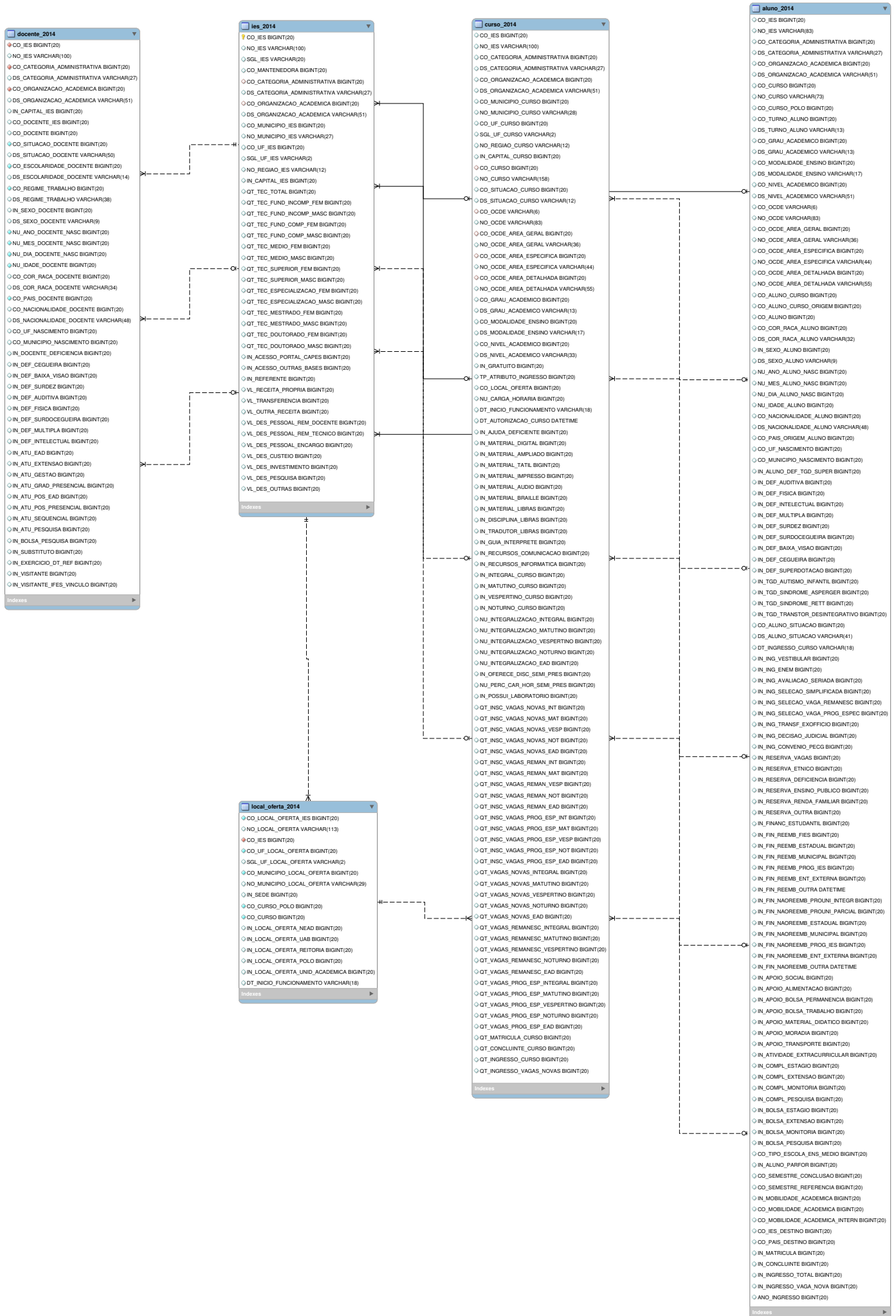
```

```
INNER JOIN DIM_IES AS IES
ON IES.CO_IES = ALUNO.CO_IES
INNER JOIN DIM_CURSO AS CURSO
ON CURSO.CO_CURSO = ALUNO.CO_CURSO
INNER JOIN DIM_TEMPO AS TEMPO
ON TEMPO.DT_INGRESSO = ALUNO.DT_INGRESSO_CURSO

INNER JOIN DIM_GEOGRAFIA AS GEOALUNO
ON GEOALUNO.COD_MUNICIPIO =
ALUNO.CO_MUNICIPIO_NASCIMENTO AND GEOALUNO.COD_UF = ALUNO.CO_UF_NASCIMENTO

INNER JOIN DIM_GEOGRAFIA_AUX AS GEOIES
ON GEOIES.COD_MUNICIPIO = IES.CO_MUNICIPIO_IES
```





# CENSO DA EDUCAÇÃO SUPERIOR 2014

## MÓDULO ALUNO

### DADOS CADASTRAIS

ID do Aluno												CPF do Aluno																							
Nome do Aluno																																			
Data de Nascimento do Aluno												Sexo do Aluno*																							
												<input type="checkbox"/> Masculino <input type="checkbox"/> Feminino																							
Nome Completo da Mãe																																			
Cor/Raça do Aluno																																			
<input type="checkbox"/> Branca <input type="checkbox"/> Preta <input type="checkbox"/> Parda <input type="checkbox"/> Amarela												<input type="checkbox"/> Indígena <input type="checkbox"/> Não dispõe da informação <input type="checkbox"/> Aluno não quis declarar a cor/raça																							
Nacionalidade												País de Origem																							
<input type="checkbox"/> Brasileira <input type="checkbox"/> Brasileira - nascido no exterior ou naturalizado <input type="checkbox"/> Estrangeira																																			
UF de Nascimento												Município de Nascimento																							
Aluno com Deficiência, Transtorno Global do Desenvolvimento ou Altas Habilidades/Superdotação																																			
<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Não dispõe da informação																																			
Tipo de Deficiência, Transtorno Global do Desenvolvimento ou Altas Habilidades/Superdotação																																			
<input type="checkbox"/> Cegueira <input type="checkbox"/> Baixa visão <input type="checkbox"/> Surdez <input type="checkbox"/> Deficiência auditiva <input type="checkbox"/> Deficiência física												<input type="checkbox"/> Surdocegueira <input type="checkbox"/> Deficiência múltipla <input type="checkbox"/> Deficiência intelectual <input type="checkbox"/> Autismo <input type="checkbox"/> Síndrome de Asperger												<input type="checkbox"/> Síndrome de Rett <input type="checkbox"/> Transtorno desintegrativo de infância <input type="checkbox"/> Altas habilidades/Superdotação											

# CENSO DA EDUCAÇÃO SUPERIOR 2014

## MÓDULO ALUNO

### VÍNCULO DO ALUNO AO CURSO



Código do Curso

Código do Polo

Turno (do Aluno no Curso)

☐ Matutino  
☐ Vespertino☐ Noturno  
☐ Integral

Situação do Vínculo do Aluno no Curso

<input type="checkbox"/> Cursando	<input type="checkbox"/> Desvinculado do curso	<input type="checkbox"/> Formado
<input type="checkbox"/> Matrícula trancada	<input type="checkbox"/> Transferido para outro curso da mesma IES	<input type="checkbox"/> Falecido

Semestre de conclusão do Curso

☐ Primeiro ☐ Segundo

Aluno PARFOR?

☐ Sim ☐ Não

Mobilidade Acadêmica

☐ Nacional ☐ Internacional

IES Destino - Mobilidade Acadêmica Nacional

Tipo de Mobilidade Acadêmica Internacional

☐ Intercâmbio ☐ Ciência sem fronteiras

País Destino - Mobilidade Acadêmica Internacional

Tipo de Escola que Concluiu o Ensino Médio

☐ Privada ☐ Pública ☐ Não dispõe de informação

Semestre de Ingresso no Curso

Forma de Ingresso/Seleção

<input type="checkbox"/> Vestibular	<input type="checkbox"/> Transferência <i>Ex-Officio</i>
<input type="checkbox"/> Enem	<input type="checkbox"/> Convênio PEC-G
<input type="checkbox"/> Avaliação seriada	<input type="checkbox"/> Decisão Judicial
<input type="checkbox"/> Seleção Simplificada (análise de currículo, entrevista, etc.)	<input type="checkbox"/> Seleção para vagas remanescentes
	<input type="checkbox"/> Seleção para vagas de programas especiais

Participa de Programa de reserva de vagas?

☐ Sim ☐ Não

Tipo de Programa de Reserva de Vagas

<input type="checkbox"/> Étnico	<input type="checkbox"/> Social/renda familiar
<input type="checkbox"/> Pessoa com deficiência	<input type="checkbox"/> Outros
<input type="checkbox"/> Estudante procedente de escola pública	

Possui Financiamento Estudantil?

☐ Sim ☐ Não

# CENSO DA EDUCAÇÃO SUPERIOR 2014

## MÓDULO ALUNO

### Tipo de Financiamento Estudantil Reembolsável

- |   |  |
|---|--|
| <input type="checkbox"/> FIES   | <input type="checkbox"/> Programa de financiamento da IES                |
| <input type="checkbox"/> Programa de financiamento do governo estadual  | <input type="checkbox"/> Programa de financiamento de entidades externas |
| <input type="checkbox"/> Programa de financiamento do governo municipal |  |

### Tipo de Financiamento Estudantil não Reembolsável

- |   |  |
|---|--|
| <input type="checkbox"/> ProUni integral                                | <input type="checkbox"/> Programa de financiamento da IES                |
| <input type="checkbox"/> ProUni parcial                                 | <input type="checkbox"/> Programa de financiamento de entidades externas |
| <input type="checkbox"/> Programa de financiamento do governo estadual  |  |
| <input type="checkbox"/> Programa de financiamento do governo municipal |  |

### Possui Apoio Social?

- ☐ Sim ☐ Não

### Tipo de Apoio Social

- |                                      |  |
|--------------------------------------|--|
| <input type="checkbox"/> Alimentação | <input type="checkbox"/> Material didático |
| <input type="checkbox"/> Moradia     | <input type="checkbox"/> Bolsa trabalho    |
| <input type="checkbox"/> Transporte  | <input type="checkbox"/> Bolsa permanência |

### Atividade Extracurricular

- ☐ Sim ☐ Não

### Tipo de Atividade Extracurricular

- |                                   |  |
|-----------------------------------|--|
| <input type="checkbox"/> Pesquisa | <input type="checkbox"/> Monitoria               |
| <input type="checkbox"/> Extensão | <input type="checkbox"/> Estágio não obrigatório |

### Possui Bolsa/Remuneração (Referente à Cada Atividade Extracurricular)?

- ☐ Sim ☐ Não



# CENSO DA EDUCAÇÃO SUPERIOR 2014

## MÓDULO CURSO

### DADOS CENSITÁRIOS

Curso Teve Aluno Vinculado em 2014?

☐ Sim ☐ Não

Motivo do Curso Sem Aluno Vinculado em 2014

☐ Curso extinto  
☐ Curso novo

☐ Curso representado por outro código de curso  
☐ Curso ativo sem demanda

Código do Curso Representado

Curso é Financiado por Convênio?

☐ Sim ☐ Não

Turno do Curso

☐ Matutino  
☐ Vespertino

☐ Noturno  
☐ Integral

Prazo Mínimo de Integralização em Anos (Por Turno ou EAD )

Vagas Novas Oferecidas  
(por Turno ou EAD)

Inscritos em Vagas Novas Oferecidas  
(por Turno ou EAD)

Vagas Remanescentes  
(por Turno ou EAD)

Inscritos em Vagas Remanescentes  
(por Turno ou EAD)

Vagas Oferecidas Para Programas Especiais  
(por Turno ou EAD)

Inscritos em Vagas Para Programas Especiais  
(por Turno ou EAD)

Curso Garante Condições de Acessibilidade às Pessoas com Deficiência?

☐ Sim ☐ Não

Recursos de Tecnologia Assistiva Disponíveis às Pessoas com Deficiência

☐ Material em Braille  
☐ Material em áudio  
☐ Recursos de informática acessível  
☐ Tradutor e intérprete de língua brasileira de sinais  
☐ Guia-Intérprete  
☐ Material didático em língua brasileira de sinais

☐ Material em formato impresso em caractere ampliado  
☐ Material pedagógico tátil  
☐ Inserção da disciplina de língua brasileira de sinais no curso  
☐ Material didático em formato impresso acessível  
☐ Recursos de acessibilidade à comunicação  
☐ Material didático digital acessível

Oferece Disciplinas Semipresenciais?

☐ Sim ☐ Não

Percentual de Carga Horária Semipresencial

Utiliza Instalações para Aulas Práticas (Laboratórios)?

☐ Sim ☐ Não

Código do Laboratório (Lista)

Tipo de Laboratório (Lista)

# CENSO DA EDUCAÇÃO SUPERIOR 2014

## MÓDULO DOCENTE

### DADOS CADASTRAIS

ID do Docente												CPF do Docente																					
<input type="text"/>												<input type="text"/>																					
Nome do Docente																																	
<input type="text"/>																																	
Data de Nascimento do Docente												Sexo do Docente																					
<input type="text"/>												<input type="checkbox"/> Masculino <input type="checkbox"/> Feminino																					
Nome Completo da Mãe																																	
<input type="text"/>																																	
Cor/Raça do Docente																																	
<input type="checkbox"/> Branca												<input type="checkbox"/> Indígena																					
<input type="checkbox"/> Preta												<input type="checkbox"/> Não dispõe da informação																					
<input type="checkbox"/> Parda												<input type="checkbox"/> Docente não quis declarar a cor/raça																					
<input type="checkbox"/> Amarela																																	
Nacionalidade												País de Origem																					
<input type="checkbox"/> Brasileira												<input type="text"/>																					
<input type="checkbox"/> Brasileira - nascido no exterior ou naturalizado												<input type="text"/>																					
<input type="checkbox"/> Estrangeira												<input type="text"/>																					
UF de Nascimento												Município de Nascimento																					
<input type="text"/>												<input type="text"/>																					
Docente com deficiências?																																	
<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Não dispõe da informação																																	
Tipo de Deficiência																																	
<input type="checkbox"/> Cegueira												<input type="checkbox"/> Deficiência auditiva												<input type="checkbox"/> Deficiência múltipla									
<input type="checkbox"/> Baixa visão												<input type="checkbox"/> Deficiência física												<input type="checkbox"/> Deficiência intelectual									
<input type="checkbox"/> Surdez												<input type="checkbox"/> Surdocegueira																					

# CENSO DA EDUCAÇÃO SUPERIOR 2014

## MÓDULO DOCENTE

### VÍNCULO DO DOCENTE A IES



#### Escolaridade

- ☐ Sem formação de nível superior  
☐ Com formação de nível superior

#### Pós-Graduação

- ☐ Especialização ☐ Doutorado  
☐ Mestrado ☐ Não possui

#### Situação do Docente na IES

- ☐ Esteve em exercício ☐ Afastado para exercício em outros órgãos/entidades  
☐ Afastado para qualificação ☐ Afastado por outros motivos  
☐ Afastado para tratamento de saúde

#### Regime de Trabalho

- ☐ Tempo integral com DE ☐ Tempo parcial  
☐ Tempo integral sem DE ☐ Horista

#### Docente Substituto

- ☐ Sim ☐ Não

#### Docente Visitante?

- ☐ Sim ☐ Não

#### Tipo de Vínculo de Docente Visitante

- ☐ Em Folha ☐ Bolsista

#### Docente em Exercício em 31/12?

- ☐ Sim ☐ Não

#### Atuação do Docente

- ☐ Ensino de pós-graduação *stricto sensu* a distância ☐ Ensino em curso sequencial de formação específica  
☐ Ensino em curso de graduação presencial ☐ Pesquisa  
☐ Ensino em curso a distância ☐ Extensão  
☐ Ensino de pós-graduação *stricto sensu* presencial ☐ Gestão, planejamento e avaliação

#### Possui Bolsa Pesquisa (Somente para Docentes com Atuação em Pesquisa)?

- ☐ Sim ☐ Não

### VÍNCULO DO Docente AO CURSO



#### Código do Curso/Área Básica ao qual o Docente está Vinculado

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--





**CENSO DA EDUCAÇÃO SUPERIOR 2014**  
**MÓDULO IES**

**ENDEREÇO DA REITORIA/SEDE ADMINISTRATIVA**[illegible]

## Dirigente Principal

[illegible]

## PROCURADOR INSTITUCIONAL

[illegible]

## LOCAL DE OFERTA

[illegible]

**CENSO DA EDUCAÇÃO SUPERIOR 2014**  
**MÓDULO IES**

Bairro do Local de Oferta												Número do Local de Oferta																																											
Complemento do Local de Oferta																																																							
CEP do Local de Oferta												UF do Local de Oferta																																											
Município do Local de Oferta																																																							
Tipo de Local de Oferta																																																							
<input type="checkbox"/>	NEAD																											<input type="checkbox"/>	Polo UAB																										
<input type="checkbox"/>	Unidade administrativa/reitoria																											<input type="checkbox"/>	Polo																										
<input type="checkbox"/>	Unidade acadêmica																																																						

## DADOS CENSITÁRIOS

Número de Funcionários Técnicos-Administrativos por Grau de Escolaridade - Feminino						Número de Funcionários Técnicos-Administrativos por Grau de Escolaridade - Masculino					
					Fundamental incompleto						Fundamental incompleto
					Fundamental completo						Fundamental completo
					Ensino médio						Ensino médio
					Ensino superior						Ensino superior
					Especialização						Especialização
					Mestrado						Mestrado
					Doutorado						Doutorado

## DADOS FINANCEIROS

Dados Financeiros Referentes à			
<input type="checkbox"/>	Mantenedora	<input type="checkbox"/>	Instituição
Receitas Auferidas			
<input type="checkbox"/>	Receitas próprias		
<input type="checkbox"/>	Transferências		
<input type="checkbox"/>	Outras receitas		
Despesas Efetuadas			
<input type="checkbox"/>	Pessoal - Remuneração de professores ativos	<input type="checkbox"/>	Outras despesas de custeio
<input type="checkbox"/>	Pessoal - Remuneração de pessoal técnico-administrativo/pedagógico	<input type="checkbox"/>	Investimentos (despesas de capital)
<input type="checkbox"/>	Pessoal - Benefícios e encargos sociais	<input type="checkbox"/>	Pesquisa e desenvolvimento
		<input type="checkbox"/>	Outras despesas



**CENSO DA EDUCAÇÃO SUPERIOR 2014**  
**MÓDULO IES**

## Acessibilidade de Conteúdo

- ☐
- Plano de aquisição gradual de acervo bibliográfico dos conteúdos básicos em formato especial

## Acessibilidade Tecnológica

- |                          |  |                          |                 |
|--------------------------|--|--------------------------|-----------------|
| <input type="checkbox"/> | Disponibiliza software de leitura para pessoas com baixa visão | <input type="checkbox"/> | Teclado virtual |
| <input type="checkbox"/> | Disponibiliza impressoras em Braille                           |                          |                 |

--	--	--	--	--	--	--	--

## INFORMAÇÕES ADICIONAIS

Possui Acesso ao Portal Capes de Periódicos?	Assina Outras Bases de Dados?
Sim	Sim
Não	Não

- ☐ Sim ☐ Não

Possui Biblioteca Digital *online* de Serviço Público?

- Possui Catálogo *online* de Serviço Público

- ☐ Sim ☐ Não

Itens do Acervo Eletrônico					

--	--	--	--	--	--

## INFRAESTRUTURA

### Quantidade de Computadores Destinados ao Uso dos Discentes

[illegible]

## Existem Condições de Acessibilidade Arquitetônica para Pessoas com Deficiência?

- |                          |     |                          |     |
|--------------------------|-----|--------------------------|-----|
| <input type="checkbox"/> | Sim | <input type="checkbox"/> | Não |
|--------------------------|-----|--------------------------|-----|

## Acessibilidade Arquitetônica ou Física

- |                          |  |                          |                                 |
|--------------------------|--|--------------------------|---------------------------------|
| <input type="checkbox"/> | Sinalização tátil  | <input type="checkbox"/> | Sinalização sonora              |
| <input type="checkbox"/> | Entrada/Saída com dimensionamento  | <input type="checkbox"/> | Sinalização visual              |
| <input type="checkbox"/> | Ambientes desobstruídos que facilitem a movimentação de cadeirantes e pessoas com deficiência visual | <input type="checkbox"/> | Banheiros adaptados             |
|                          |  | <input type="checkbox"/> | Espaço de atendimento adaptados |
|                          |  | <input type="checkbox"/> | Mobiliário adaptado             |
|                          |  | <input type="checkbox"/> | Bebedouros e lavabos adaptados  |

Equipamento eletromecânico (elevador, esteira rolante, etc)

- ☐ Sim ☐ Não ☐ Não se aplica

Rampa de acesso com corrimão

- ☐
- Sim
- ☐
- Não
- ☐
- Não se aplica

Infraestrutura Polo EAD

## Informa as Tecnologias e Equipamentos do Polo

- |                          |                                   |                          |  |
|--------------------------|-----------------------------------|--------------------------|--|
| <input type="checkbox"/> | Equipamento para videoconferência | <input type="checkbox"/> | Conexão à internet banda larga                   |
| <input type="checkbox"/> | Sala de coordenação do polo       | <input type="checkbox"/> | Salas equipadas para o atendimento pelos tutores |
| <input type="checkbox"/> | Microcomputadores                 |                          |  |

# CENSO DA EDUCAÇÃO SUPERIOR 2014

## MÓDULO IES

### Infraestrutura Local de Oferta Presencial

#### Instalações da Unidade

- |   |  |
|---|--|
| <input type="checkbox"/> Restaurante universitário                          | <input type="checkbox"/> Cantina/Lanchonete                          |
| <input type="checkbox"/> Quadra poliesportiva                               | <input type="checkbox"/> Pista de atletismo                          |
| <input type="checkbox"/> Piscina olímpica/semi-olímpica                     | <input type="checkbox"/> Campo para práticas esportivas              |
| <input type="checkbox"/> Quadra coberta/ginásio                             | <input type="checkbox"/> Auditório/Teatro                            |
| <input type="checkbox"/> Cinema   | <input type="checkbox"/> Redes sem fio para comunidade acadêmica     |
| <input type="checkbox"/> Vestiário  | <input type="checkbox"/> Posto de atendimento para primeiro socorros |
| <input type="checkbox"/> Serviços   | <input type="checkbox"/> Bicicletário                                |
| <input type="checkbox"/> Estacionamento para docente/técnico administrativo | <input type="checkbox"/> Estacionamento para aluno                   |
|   | <input type="checkbox"/> Moradia estudantil                          |